

# Interference Estimation with Applications to Blind Multiple-Access Communication Over Fading Channels

Michael L. McCloud, *Student Member, IEEE*, and Louis L. Scharf, *Fellow, IEEE*

**Abstract**—We consider the detection of nonorthogonal multipulse signals on multiple-access fading channels. The generalized maximum-likelihood rule is employed to decode users whose complex fading gains are unknown. We develop geometrical interpretations for the resulting detectors and their corresponding asymptotic efficiencies. The generalized maximum-likelihood detection rule is then applied to find a matched subspace detector for the frequency-selective fading channel, under the assumption of a short coherence time (or long coherence time without the computational power to track the fading parameters). We propose blind implementations of these detectors for nonorthogonal multipulse signaling on both frequency-nonselective and frequency-selective multiple-access fading channels. These blind detectors extend the results of Wang and Poor to multipulse modulation and fast frequency selective fading. For comparison, the minimum mean-squared error decision rules for these channels are derived and blind implementations of their corresponding detectors are developed.

**Index Terms**—Blind detection, fading channels, generalized likelihood functions, interference identification, multiuser detection.

## I. INTRODUCTION

OW-complexity detectors are of fundamental importance in multiple-user communication channels. Linear complexity detectors are termed decentralized detectors since they make decisions on each user of a multiple-access channel independently. An important decentralized detector for linear modulation is the decorrelating detector derived through the generalized likelihood principle by Lupas and Verdú [1]. This detector has been generalized to nonorthogonal multipulse modulation (NMM) on both coherent and noncoherent channels in [2]–[6]. The resulting detectors have a rich geometric structure, choosing the signal which makes the smallest angle with the measurement, in a subspace orthogonal to that occupied by the competing users.

Recently, a subspace technique has been proposed by Wang and Poor for implementing blind, decentralized, decorrelating detectors for linear signaling, without explicit knowledge of the interference subspace [7]. Implicit in this technique is an algorithm for identifying the interference subspace and the projection matrix onto its complement. In Section VII of this paper we

explicitly state and generalize this subspace identification algorithm to the higher rank signal subspaces required for NMM signaling over fading channels. The estimator is shown to be strongly consistent and asymptotic bounds on its performance are obtained.

In Section III of this paper we review the generalized maximum-likelihood (GML) detectors derived in [3]–[6] for NMM. Geometric interpretations of these detectors are presented and their invariances are noted. The detectors are derived for coherent and noncoherent signaling with an unknown gain term corresponding to *frequency nonselective* fading and for multi-rank signaling induced through *frequency selective* fading. For both problems we assume that the fading parameters are varying sufficiently fast so that reliable estimation is not practical for contiguous symbol periods, although they are assumed constant over at least one signaling period. In each case, we assume that the signaling is synchronous and we derive detectors that are *invariant* to the fading through the generalized maximum-likelihood (GML) decision rule. In Section VIII we apply the interference identification results of Section VII to the GML detectors of Section III, developing *blind* detectors that adapt to the unknown multiple-access interference (MAI). The computational complexity of the blind detectors is examined. It is shown that the detectors are feasible to implement even at relatively high data rates due to the relatively small computational burden of the blind detectors when fast subspace tracking is employed.

Finally, in Section X we give a brief derivation of the minimum mean-squared error (MMSE) detectors for NMM signaling for both the coherent and the noncoherent channel. Blind implementations of the detectors are presented and a numerical example is presented to compare the MMSE and GML detectors for the Rayleigh fading channel. For this example, the detectors are seen to have essentially identical performance, with the GML detector performing slightly better asymptotically. It is shown that, unlike the case of linear modulation, the MMSE detector is not necessarily the limit of the generalized likelihood detector. This topic is discussed in more detail in [31], in which the authors derive an asymptotic performance bound on the MMSE detector. This bound generally differs from that of the GML detector.

## II. NOTATION AND LINEAR ALGEBRAIC RESULTS

We will denote matrices and vectors with bold face type, using capital letters for matrices and lower case letters for vectors. Given a matrix  $\mathbf{A} \in \mathbb{C}^{N \times M}$  we denote the linear subspace

Manuscript received December 1, 1998; revised July 1, 1999. This work was supported by the National Science Foundation under Contract ECS 9979400 and by the Office of Naval Research under Contracts N00014-89-J-1070 and N0014-00-1-0033.

The authors are with the Department of Electrical Engineering, University of Colorado, Boulder, CO 80303 USA.

Communicated by M. L. Honig, Associate Editor for Communications.

Publisher Item Identifier S 0018-9448(00)02888-1.

of  $\mathbb{C}^N$  spanned by the columns of  $\mathbf{A}$  by  $\langle \mathbf{A} \rangle$ . We use the symbol  $\mathbf{P}_A$  to denote the orthogonal projection matrix with range space  $\langle \mathbf{A} \rangle$ , and when  $\mathbf{A}$  has full column rank then the projection onto the subspace  $\langle \mathbf{A} \rangle$  is

$$\mathbf{P}_A = \mathbf{A}(\mathbf{A}^* \mathbf{A})^{-1} \mathbf{A}^* = \mathbf{P}_A^* = \mathbf{P}_A^2. \quad (1)$$

We denote the subspace of  $\mathbb{C}^N$  orthogonal to  $\langle \mathbf{A} \rangle$  by  $\langle \mathbf{A} \rangle^\perp$ , with corresponding projection  $\mathbf{P}_A^\perp = \mathbf{I} - \mathbf{P}_A$ .

When  $[\mathbf{A}, \mathbf{B}]$  has full column rank we may form the orthogonal projection onto the subspace  $\langle [\mathbf{A}, \mathbf{B}] \rangle$  as

$$\mathbf{P}_{AB} = [\mathbf{A}, \mathbf{B}][\mathbf{A}, \mathbf{B}]^* [\mathbf{A}, \mathbf{B}]^{-1} [\mathbf{A}, \mathbf{B}]^*. \quad (2)$$

This orthogonal projection may be rewritten as the sum of two orthogonal projections, or as the sum of two *oblique* projections [8]:

$$\mathbf{P}_{AB} = \mathbf{P}_B + \mathbf{P}_{\mathbf{P}_B^\perp \mathbf{A}} = \mathbf{E}_{AB} + \mathbf{E}_{BA}. \quad (3)$$

The oblique projection matrix  $\mathbf{E}_{AB}$  has range space  $\langle \mathbf{A} \rangle$  and null space  $\langle \mathbf{B} \rangle$ . This matrix is idempotent but not symmetric and satisfies  $\mathbf{E}_{AB} \mathbf{A} = \mathbf{A}$  and  $\mathbf{E}_{AB} \mathbf{B} = \mathbf{0}$ . It has the algebraic representation [8]

$$\mathbf{E}_{AB} = \mathbf{A} (\mathbf{A}^* \mathbf{P}_B^\perp \mathbf{A})^{-1} \mathbf{A}^* \mathbf{P}_B^\perp. \quad (3a)$$

Note that  $\mathbf{E}_{AB} = \mathbf{P}_A$  when  $\langle \mathbf{B} \rangle = \langle \mathbf{A} \rangle^\perp$ . These properties may be easily extended to more finely resolve the projection  $\mathbf{P}_S$ , as is shown in the following lemma.

**Lemma II.1:** Given a matrix  $\mathbf{S} = [\mathbf{s}_1 \ \mathbf{s}_2 \ \cdots \ \mathbf{s}_M]$  such that  $\mathbf{S} \in \mathbb{C}^{N \times M}$  and  $\mathbf{S}^* \mathbf{S} > \mathbf{0}$ , then we may decompose the orthogonal projection  $\mathbf{P}_S$  as

$$\mathbf{P}_S = \mathbf{S}(\mathbf{S}^* \mathbf{S})^{-1} \mathbf{S}^* = \sum_{i=1}^M \mathbf{E}_{\mathbf{s}_i, \mathbf{S}_i}.$$

Here we have defined  $\mathbf{S}_i$  as the matrix  $\mathbf{S}$  with its  $i$ th column,  $\mathbf{s}_i$ , removed. The oblique projection matrix  $\mathbf{E}_{\mathbf{s}_i, \mathbf{S}_i}$  is computed as in (3a).

*Proof:* See Appendix A.  $\square$

**Theorem II.2:** Given a full rank matrix  $\mathbf{S} \in \mathbb{C}^{N \times M}$ , the Moore–Penrose pseudo-inverse of  $\mathbf{S}$  is given by

$$\mathbf{S}^+ = \begin{bmatrix} \mathbf{s}_1^* \mathbf{P}_{\mathbf{S}_1}^\perp / \left( \mathbf{s}_1^* \mathbf{P}_{\mathbf{S}_1}^\perp \mathbf{s}_1 \right) \\ \mathbf{s}_2^* \mathbf{P}_{\mathbf{S}_2}^\perp / \left( \mathbf{s}_2^* \mathbf{P}_{\mathbf{S}_2}^\perp \mathbf{s}_2 \right) \\ \vdots \\ \mathbf{s}_M^* \mathbf{P}_{\mathbf{S}_M}^\perp / \left( \mathbf{s}_M^* \mathbf{P}_{\mathbf{S}_M}^\perp \mathbf{s}_M \right) \end{bmatrix} \quad (4)$$

with  $\mathbf{S}^+ \mathbf{S} = \mathbf{I}$  and  $\mathbf{S} \mathbf{S}^+ = \mathbf{P}_S$ .

*Proof:* This is readily verified by expanding  $\mathbf{S} \mathbf{S}^+$  and  $\mathbf{S}^+ \mathbf{S}$ .  $\square$

The distance between the subspaces  $\langle \mathbf{A} \rangle$  and  $\langle \mathbf{B} \rangle$  is defined to be [9]

$$\text{dist}(\langle \mathbf{A} \rangle, \langle \mathbf{B} \rangle) = \|\mathbf{P}_A - \mathbf{P}_B\|_2 = \max |\sin(\theta_i)| \quad (5)$$

where  $\{\theta_i\}$  are the principle angles between the subspaces  $\langle \mathbf{A} \rangle$  and  $\langle \mathbf{B} \rangle$ . Here the 2-norm of the matrix  $\mathbf{A}$ , denoted  $\|\mathbf{A}\|_2$ , is the largest singular value of  $\mathbf{A}$ .

We will also deal with vectors  $\mathbf{h} \in \mathbb{L}^2$ , i.e., continuous-time, finite-energy signals. We will use the usual inner product for these vectors, namely,

$$\langle \mathbf{h}, \mathbf{u} \rangle = \int_{t=-\infty}^{\infty} h(t) u^*(t) dt. \quad (6)$$

Given an indexed set of signals  $\{\mathbf{u}(n)\}$ ,  $\mathbf{u}(n) \in \mathbb{L}^2$ , let the vector  $\mathbf{U} \in \{\mathbb{L}^2\}^N$  be formed by stacking the  $\mathbf{u}(n)$ 's; that is,  $\mathbf{U} = [\mathbf{u}(1), \dots, \mathbf{u}(N)]^T$ . We then find the *multilinear* inner product taking  $\{\mathbb{L}^2\}^N \times \mathbb{L}^2$  into  $\mathbb{C}^N$  to be

$$\{\langle \mathbf{h}, \mathbf{U} \rangle\}_i = \langle \mathbf{h}, \mathbf{u}(i) \rangle = \int_{t=-\infty}^{\infty} h(t) u_i^*(t) dt \quad (7)$$

where  $\{\}_i$  denotes the  $i$ th element.

### III. MULTIUSER COMMUNICATIONS WITH NONORTHOGONAL MULTIPULSE MODULATION ON THE FREQUENCY NONSELECTIVE FADING CHANNEL

We now apply our linear algebraic results to generalized maximum-likelihood detectors for multiuser communication with NMM. The signaling model that we assume is quite general, including as special cases the linear modulation schemes such as phase-shift keying and quadrature-amplitude modulation, as well as the “nonlinear” (linear in a higher-dimensional basis set) modulation schemes such as frequency-shift keying. We examine frequency-nonspecific fading (unknown gain) in this section and frequency-selective fading in Section VI. For the nonselective fading channel, we examine both coherent and noncoherent signaling. In all cases we assume synchronous signaling and decentralized detectors. This ensures linear complexity at the base station for demodulation on each baud interval.

#### A. Coherent Signaling Over the Frequency Nonselective Fading Channel

Let  $K$  users share a common communication channel. Each user employs synchronous coherent NMM signaling with an unknown gain affecting each user. Specifically, each user transmits one of the signals from a personal  $M$ -ary signal constellation at each baud period. Let the received continuous time signal be

$$\mathbf{r} = \sum_{k=1}^K \mu(k) \mathbf{s}(k) + \mathbf{n} \quad (8)$$

where  $\mathbf{s}(k) \in \mathfrak{S}(k) = \{\mathbf{s}_m(k)\}_{m=1}^M$  is the signal transmitted by the  $k$ th user on the  $T$ -second baud interval,  $\mu(k)$  is the  $k$ th users' fading parameter (assumed constant across a signaling period), and  $\mathbf{n}$  is an additive white Gaussian noise process. Let the signals  $\cup_{k=1}^K \mathfrak{S}(k)$  have an orthonormal basis  $\{\mathbf{u}(n)\}_{n=1}^N$  in  $\mathbb{L}^2$  with  $N \leq KM$ . For the detection problem we form a vector sufficient statistic  $\mathbf{y} = G(\mathbf{r})$ ,  $G: \mathbb{L}^2 \rightarrow \mathbb{R}^N$ , by correlating  $\mathbf{r}$  against each basis function; that is,

$$\mathbf{y} = G(\mathbf{r}) = \langle \mathbf{r}, \mathbf{U} \rangle \quad (9)$$

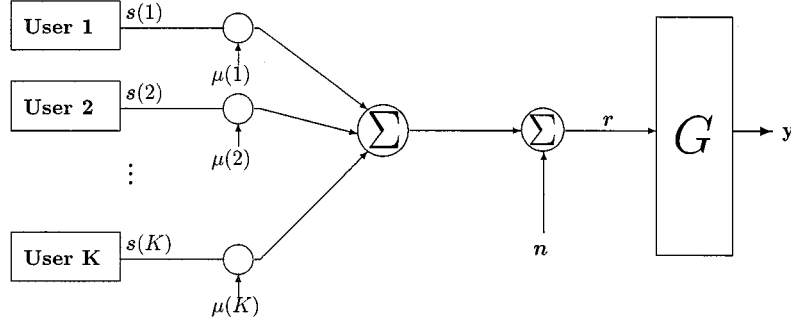


Fig. 1. System model for frequency-nonsselective fading.

$$= \mathcal{H} \mathcal{D} \mathbf{b} + \mathbf{n} \quad (10)$$

where the multilinear inner product  $\langle \mathbf{r}, \mathbf{U} \rangle$  is defined in (7). The matrix  $\mathcal{H} = [\mathbf{H}(1), \mathbf{H}(2), \dots, \mathbf{H}(K)]$  contains the signal vectors for each user with  $\mathbf{H}(k) = [\mathbf{h}_1(k), \mathbf{h}_2(k), \dots, \mathbf{h}_M(k)]$  and

$$\mathbf{h}_m(k) = \langle \mathbf{s}_m(k), \mathbf{U} \rangle. \quad (11)$$

The matrix

$$\mathcal{D} = \text{diag}\{\mu(1)\mathbf{I}, \mu(2)\mathbf{I}, \dots, \mu(K)\mathbf{I}\}$$

is a diagonal  $MK \times MK$  matrix of the users' fading coefficients and

$$\mathbf{b} = [\mathbf{b}^T(1), \mathbf{b}^T(2), \dots, \mathbf{b}^T(K)]^T$$

is an  $MK \times 1$  vector. Each  $\mathbf{b}(k)$  is a column of the  $M \times M$  identity matrix that selects the signal transmitted by user  $k$ ,  $\mathbf{H}(k)\mathbf{b}_m(k) = \mathbf{h}_m(k)$ . Assuming equally likely signals for each user and deterministic gain coefficients, the measurement has first- and second-order statistics

$$\mathbf{E}[\mathbf{y}] = \frac{1}{M} \mathcal{H} \mathcal{D} \mathbf{1} \text{ and } \mathbf{E}[\mathbf{y} \mathbf{y}^T] = \mathcal{H} \mathcal{D} \mathbf{R}_{\mathbf{b}} \mathcal{D}^T \mathcal{H}^T + \sigma^2 \mathbf{I} \quad (12)$$

where

$$\mathbf{R}_{\mathbf{b}} = \mathbf{E}[\mathbf{b} \mathbf{b}^T] = \begin{bmatrix} \frac{1}{M} \mathbf{I} & \frac{1}{M^2} \mathbf{Q} & \cdots & \frac{1}{M^2} \mathbf{Q} \\ \frac{1}{M^2} \mathbf{Q} & \frac{1}{M} \mathbf{I} & \cdots & \frac{1}{M^2} \mathbf{Q} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{M^2} \mathbf{Q} & \frac{1}{M^2} \mathbf{Q} & \cdots & \frac{1}{M} \mathbf{I} \end{bmatrix} \quad (13)$$

and  $\mathbf{Q} = \mathbf{1} \mathbf{1}^T$  with  $\mathbf{1} = [1 \ 1 \ \dots \ 1]^T$ .

We may expand this model for when the  $k$ th user is of interest and has transmitted the  $m$ th signal from  $\mathfrak{S}(k)$  as

$$\mathbf{y} = \mu(k) \mathbf{h}_m(k) + \mathbf{S}(k) \boldsymbol{\beta} + \mathbf{n}. \quad (14)$$

The vector  $\mathbf{S}(k) \boldsymbol{\beta}$  is the multiple-access interference corresponding to users  $l \neq k$ . The system model for this problem is shown in Fig. 1. We assume that the users communicate independently.

This is a general model for code-division multiple access (CDMA) communications, but to distinguish it from direct-sequence CDMA (DS-CDMA), Varanasi and Guess [10] have proposed that it be called Correlated-Waveform Multiple Access (CWMA). The model is general enough to include direct-sequence Code-Division Multiple Access (DS-CDMA) tech-

niques wherein the  $N$  basis functions can be chosen as time delays of a common chip pulse  $\psi(t)$  with chip period  $T_C$ ; that is,  $u_n(t) = \psi(t - nT_C)$ . It also models Time-Division Multiple Access (TDMA) wherein the basis functions are again time delays of the same basic pulse and the signaling vectors  $\mathbf{h}_m(k)$  are constrained to have only one nonzero entry (in the  $k$ th position). The coherent version also models (block) coded multiple-access communications wherein each symbol vector  $\mathbf{h}_m(k)$  is a codeword representing an  $M$ -bit information block. Similarly, the noncoherent version of this model, discussed below, is general enough to include such symboling schemes as frequency shift keying, wherein the basis functions are truncated complex exponentials of different frequencies. Notice that for the noncoherent channel, our detector and performance analysis allow for correlated signaling. This allows for bandwidth-efficient signaling in conjunction with noncoherent detection.

#### B. Noncoherent Signaling Over the Frequency Nonselective Fading Channel

When the channel is noncoherent we employ essentially the same model, with  $\mu(k)$  now representing an unknown complex gain and  $\mathbf{r}$  replaced by its complex baseband representation. The generalized sampler,  $G$  now maps  $\mathbb{L}^2$  into  $\mathbb{C}^N$ . We have the model

$$\mathbf{y} = \mathcal{H} \mathcal{D} \mathbf{b} + \mathbf{n} \quad (15)$$

where  $\mathcal{H}$  and  $\mathbf{b}$  are defined as in the coherent case but the fading coefficient matrix is now given by

$$\mathcal{D} = \text{diag}\{\mu_1(1), \mu_2(1), \dots, \mu_M(1), \dots, \mu_1(K), \dots, \mu_M(K)\}.$$

The individual fading parameters are modeled as having amplitude  $|\mu(k)|$ , independent of the transmitted symbol, but a phase which may be hypothesis-dependent. Such models arise when the fading process is assumed to have a flat magnitude response but possibly a frequency-selective phase. Modeling the gain coefficients as zero-mean random variables, the measurement  $\mathbf{y}$  has first- and second-order statistics

$$\mathbf{E}[\mathbf{y}] = \mathbf{0} \text{ and } \mathbf{E}[\mathbf{y} \mathbf{y}^*] = \mathcal{H}^* \mathbf{F} \mathcal{H} + \sigma^2 \mathbf{I} \quad (16)$$

where

$$\mathbf{F} = \text{diag}\{\mathbf{E}[|\mu(1)|^2] \mathbf{I}, \dots, \mathbf{E}[|\mu(K)|^2] \mathbf{I}\}$$

(assuming  $\mathbf{E}[e^{j\theta_m(k)} e^{-j\theta_n(k)}] = \delta_{m,n}$ ).

When the  $m$ th signal is transmitted by user  $k$ , the measurement model becomes

$$\mathbf{y} = |\mu(k)|e^{j\phi_m(k)}\mathbf{h}_m(k) + \mathbf{S}(k)\boldsymbol{\beta} + \mathbf{n}. \quad (17)$$

The vector  $\mathbf{S}\boldsymbol{\beta}$  accounts for MAI.

#### IV. GENERALIZED MAXIMUM-LIKELIHOOD DETECTION

We employ the generalized maximum-likelihood (GML) detection rule for the coherent and the noncoherent channels. The idea behind this detection strategy is that if the gain  $\mu(k)$  and the multiple-access interference term,  $\mathbf{S}(k)\boldsymbol{\beta}$  were known to the receiver, the optimum (minimum probability of error) decision rule would be

$$\hat{m}(k) = \arg \min_m \|\mathbf{y} - \mu(k)\mathbf{h}_m(k) - \mathbf{S}(k)\boldsymbol{\beta}\|^2 \quad (18)$$

assuming equal priors. In the absence of this information we form maximum-likelihood estimates of  $\mu(k)$  and  $\mathbf{S}\boldsymbol{\beta}$  and substitute those into the detector. In the following two sections we give a brief outline of this procedure for the coherent and noncoherent channels. The noncoherent GML rule is also derived in [2]–[6].

##### A. The Coherent GML Detector

When the channel is phase-coherent we may find the maximum-likelihood estimates of  $\mu(k)\mathbf{h}_m(k)$  and  $\mathbf{S}(k)\boldsymbol{\beta}$  as [11], [3]

$$\begin{aligned} \hat{\mu}_m(k)\mathbf{h}_m(k) &= \begin{cases} \mathbf{E}_{\mathbf{h}_m(k)\mathbf{S}(k)}\mathbf{y}, & \mathbf{h}_m^T(k)\mathbf{P}_{\mathbf{S}(k)}^\perp\mathbf{y} \geq 0 \\ \mathbf{0}, & \mathbf{h}_m^T(k)\mathbf{P}_{\mathbf{S}(k)}^\perp\mathbf{y} < 0 \end{cases} \\ \mathbf{S}(k)\hat{\boldsymbol{\beta}}_m(k) &= \mathbf{E}_{\mathbf{S}(k)\mathbf{h}_m(k)}\mathbf{y} \end{aligned} \quad (19)$$

where  $\mathbf{E}_{\mathbf{S}(k)\mathbf{h}_m(k)}$  and  $\mathbf{E}_{\mathbf{h}_m(k)\mathbf{S}(k)}$  are the oblique projection matrices defined in Section II. Notice that we have included the subscript on the estimates to indicate that  $\hat{\mu}_m(k)$  is the maximum-likelihood estimate of the  $k$ th user's gain under the  $m$ th hypothesis, and similarly for  $\hat{\boldsymbol{\beta}}_m(k)$ .

The resulting GML detection rule is

$$\begin{aligned} \hat{m}(k) &= \arg \min_m \|\mathbf{y} - \hat{\mu}_m(k)\mathbf{h}_m(k) - \mathbf{S}(k)\hat{\boldsymbol{\beta}}_m(k)\|^2 \\ &= \arg \max_m \frac{\mathbf{y}^T \mathbf{P}_{\mathbf{S}(k)}^\perp \mathbf{h}_m(k)}{\left(\mathbf{h}_m^T(k) \mathbf{P}_{\mathbf{S}(k)}^\perp \mathbf{h}_m(k)\right)^{1/2}}. \end{aligned} \quad (20)$$

Notice that in the case of *linear* modulation,  $\mathbf{h}_m(k) = \alpha_m \mathbf{h}(k)$ , with  $\mathbf{h}(k)$  a fixed vector and  $\alpha_m$  a scalar, the GML detector is the decorrelating detector of Lupas and Verdú [1], which was derived from a GML framework.

##### B. The Noncoherent GML Detector

When the channel is noncoherent we proceed in the same manner, with the maximum-likelihood estimates now given by

$$\begin{aligned} \hat{\mu}_m(k)\mathbf{h}_m(k) &= \mathbf{E}_{\mathbf{h}_m(k)\mathbf{S}(k)}\mathbf{y}, \\ \mathbf{S}(k)\hat{\boldsymbol{\beta}}_m(k) &= \mathbf{E}_{\mathbf{S}(k)\mathbf{h}_m(k)}\mathbf{y}. \end{aligned} \quad (21)$$

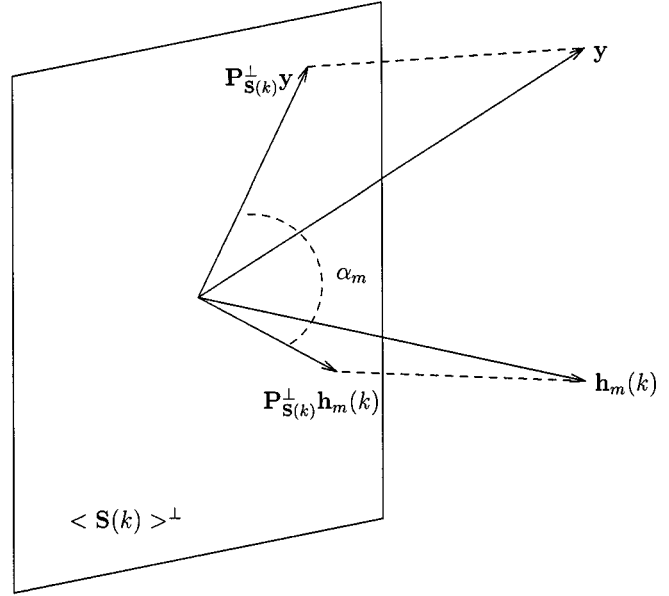


Fig. 2. Geometrical interpretation of coherent detector for the  $k$ th user. The detector makes the decision  $\hat{m}(k) = \arg \max_m \alpha_m$  where  $\alpha_m$  is the angle between  $\mathbf{y}$  and  $\mathbf{h}_m(k)$  in the interference-free subspace  $\langle \mathbf{S}(k) \rangle^\perp$ .

The corresponding noncoherent GML detection rule is

$$\hat{m}(k) = \arg \max_m \frac{|\mathbf{y}^* \mathbf{P}_{\mathbf{S}(k)}^\perp \mathbf{h}_m(k)|^2}{\left(\mathbf{h}_m^*(k) \mathbf{P}_{\mathbf{S}(k)}^\perp \mathbf{h}_m(k)\right)}. \quad (22)$$

##### C. Geometry of the GML Detectors

Defining the “interference-free signal matrix”  $\mathbf{G} = \mathbf{P}_{\mathbf{S}(k)}^\perp \mathbf{H}(k)$ , we notice that

$$\mathbf{P}_{\mathbf{S}(k)}^\perp \mathbf{h}_m(k) = \mathbf{P}_{\mathbf{G}} \mathbf{h}_m(k).$$

Thus we can replace the matrix  $\mathbf{P}_{\mathbf{S}(k)}^\perp$  by  $\mathbf{P}_{\mathbf{G}}$  everywhere it occurs in (21) and (22). Then we may normalize each detector by the measurement's energy in the space  $\langle \mathbf{G} \rangle$  to get the following geometrical interpretation [3], [5] of the GML rules:<sup>1</sup>

$$\hat{m}(k) = \begin{cases} \arg \max_m \cos(\mathbf{y}, \mathbf{h}_m(k); \mathbf{P}_{\mathbf{G}}), & \text{coherent} \\ \arg \max_m \cos^2(\mathbf{y}, \mathbf{h}_m(k); \mathbf{P}_{\mathbf{G}}), & \text{noncoherent.} \end{cases} \quad (23)$$

We see that the detector chooses the signal  $\mathbf{h}_m(k)$  which makes the smallest angle to the measurement in the perpendicular subspace  $\langle \mathbf{S} \rangle^\perp$ . Fig. 2 illustrates this geometrical interpretation of the detector. It is also worth noting the invariances of the two detectors. The coherent detector is invariant to positive scaling and to addition of vectors lying in  $\langle \mathbf{S}(k) \rangle$ . The noncoherent detector is invariant to gain and rotations (by a complex exponential) as

<sup>1</sup>When all vectors lie in  $\mathbb{R}^N$  we define

$$\cos(\mathbf{x}, \mathbf{y}; \mathbf{A}) = \langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{A}} / (\langle \mathbf{x}, \mathbf{x} \rangle_{\mathbf{A}} \langle \mathbf{y}, \mathbf{y} \rangle_{\mathbf{A}})^{1/2}$$

and use the notation  $\cos(\mathbf{x}, \mathbf{y})$  when  $\mathbf{A} = \mathbf{I}$ . When we are in  $\mathbb{C}^N$  we define

$$\cos^2(\mathbf{x}, \mathbf{y}; \mathbf{A}) = |\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{A}}|^2 / (\langle \mathbf{x}, \mathbf{x} \rangle_{\mathbf{A}} \langle \mathbf{y}, \mathbf{y} \rangle_{\mathbf{A}}).$$

well as to addition of vectors in  $\langle \mathbf{S}(k) \rangle$ . These generalize the invariance sets of the subspace detectors presented in [12] to the multiple hypothesis testing problem.

## V. ASYMPTOTIC EFFICIENCY OF THE GML DETECTORS

A common performance measure for multiuser detection is the Multiuser Asymptotic Efficiency (MAE), which measures the performance loss of a detection strategy in multiple-access interference relative to the case of a single-user channel. The MAE for user  $k$  is defined by [1]

$$\eta(k) = \sup \left\{ 0 \leq r \leq 1, \lim_{\sigma^2 \rightarrow 0} \frac{P_\phi(\sigma^2)}{P_{\text{SU}}(\sigma^2 r^{-1})} < \infty \right\} \quad (24)$$

where  $P_\phi(\sigma^2)$  is the probability of error for the  $k$ th user employing the GML detector  $\phi$  with additive white Gaussian noise (AWGN) power  $\sigma^2$ , and  $P_{\text{SU}}(\sigma^2 r^{-1})$  is the probability of error for the GML decentralized detector in the absence of interfering users ( $\mathbf{S} = \mathbf{0}$ ) with effective noise power  $\sigma^2/r$ .

We find the MAE of the detectors using the upper and lower bounds as in [13]. Suppose we can find functions

$$0 < \underline{P}_{\text{SU}}(\sigma^2 r^{-1}) \leq P_{\text{SU}}(\sigma^2 r^{-1})$$

and

$$P_\phi(\sigma^2) \leq \bar{P}_\phi(\sigma^2) < 1.$$

Then for any pair  $(\sigma^2, r)$  we have

$$\frac{P_\phi(\sigma^2)}{P_{\text{SU}}(\sigma^2 r^{-1})} \leq \frac{\bar{P}_\phi(\sigma^2)}{\underline{P}_{\text{SU}}(\sigma^2 r^{-1})} \quad (25)$$

and so

$$\eta(k) \leq \bar{\eta}(k) = \sup_{0 \leq r \leq 1} \lim_{\sigma^2 \rightarrow 0} \frac{\bar{P}_\phi(\sigma^2)}{\underline{P}_{\text{SU}}(\sigma^2 r^{-1})}. \quad (26)$$

We can generate a lower bound  $\eta(k) \leq \eta(k)$  in an analogous fashion using the bounding functions  $\underline{P}_\phi(\sigma^2)$  and  $\bar{P}_{\text{SU}}(\sigma^2 r^{-1})$ . Clearly, whenever  $\bar{\eta}(k) = \eta(k)$  we have an exact expression for  $\eta(k)$ , namely,  $\eta(k) = \bar{\eta}(k)$ . In the next section we use a union upper bound and a two-signal lower bound on the probability of error to derive the asymptotic efficiency of each of the detectors. This implies that we only need an expression for the probability of a binary error, choosing hypothesis  $n$  when hypothesis  $m$  is in effect, for the detectors. This process is sketched in the next two sections and the resulting values of the MAE are given.

### A. Asymptotic Efficiency of the Coherent GML Detector

To form our bounds for the probability of error using the coherent GML detector in (23) we begin with the binary error probability. We will find an expression for the single-user detector (in the absence of multiple-access interference). The multiple-access case is treated in an identical fashion with  $\mathbf{h}_m(k)$  replaced by  $\mathbf{P}_{\mathbf{S}(k)}^\perp \mathbf{h}_m(k)$ .

The probability of choosing signal  $n$  when signal  $m$  was transmitted using the decision rule in (23) is

$$P \left[ \frac{\mathbf{h}_m^T(k) \mathbf{y}}{(\mathbf{h}_m^T(k) \mathbf{h}_m(k))^{1/2}} < \frac{\mathbf{h}_n^T(k) \mathbf{y}}{(\mathbf{h}_n^T(k) \mathbf{h}_n(k))^{1/2}} \mid \mathbf{h}_m(k) \right] \\ = Q \left( \frac{\mu(k) (\mathbf{h}_m^T(k) \mathbf{h}_m(k))^{1/2}}{\sigma \sqrt{2}} [1 - \cos(\mathbf{h}_m(k), \mathbf{h}_n(k))]^{1/2} \right).$$

This allows us to form the two-signal lower bound

$$\underline{P}_{\text{SU}}(\sigma^2 r^{-1}) = \frac{1}{M} Q \left( \min_{m \neq n} \frac{\mu(k) (\mathbf{h}_m^T(k) \mathbf{h}_m(k))^{1/2}}{\sigma \sqrt{2}} \times [1 - \cos(\mathbf{h}_m(k), \mathbf{h}_n(k))]^{1/2} \right). \quad (27)$$

We can find expressions for  $\bar{P}_\phi(\sigma^2)$ ,  $\underline{P}_\phi(\sigma^2)$ , and  $\bar{P}_{\text{SU}}(\sigma^2 r^{-1})$  in a similar manner. Substituting these values into (26) we find the asymptotic efficiency of our detector to be given by (28) at the bottom of this page. We see that the asymptotic efficiency is nonzero as long as the user's signal set is not contained in the span of the interfering user's signals, and no two signals are positive multiples of each other.

### B. Asymptotic Efficiency of the Noncoherent Detector

In order to derive an expression for the MAE of the noncoherent detector, we start as in Section V-A and find the probability of a binary error with our detector. An independent derivation of this result was presented in a somewhat different notation in [4] and [6]. The probability of choosing signal  $\mathbf{h}_n(k)$  when  $\mathbf{h}_m(k)$  was transmitted is given by

$$P_e = P[(\mathbf{y}^* \mathbf{P}_n \mathbf{y} - \mathbf{y}^* \mathbf{P}_m \mathbf{y}) > 0 \mid \mathbf{h}_m(k)] \\ = P[\mathbf{y}^* \Delta \mathbf{P} \mathbf{y} > 0 \mid \mathbf{h}_m(k)]$$

where we have defined (29) and (30) at the top of the following page.

*Lemma V.1:* Let  $\mathbf{P}_1 = \mathbf{U}_1 \mathbf{U}_1^*$  and  $\mathbf{P}_2 = \mathbf{U}_2 \mathbf{U}_2^*$  be two orthogonal  $N \times N$  projection matrices with range spaces of the same dimension  $r$ . Then the matrix  $\Delta \mathbf{P} = \mathbf{P}_1 - \mathbf{P}_2$  has  $N - 2r$  eigenvalues equal to zero and  $2r$  eigenvalues given by  $\pm \sin(\theta_i)$ , where  $\theta_i$  is the  $i$ th principle angle between the subspaces  $\langle \mathbf{P}_1 \rangle$  and  $\langle \mathbf{P}_2 \rangle$ .

*Proof:* The characteristic polynomial for  $\Delta \mathbf{P}$  is given by

$$p(s) = \det(s\mathbf{I} - \mathbf{P}_1 + \mathbf{P}_2) \\ = \det(s\mathbf{I} + \mathbf{P}_1) \det(\mathbf{I} - \mathbf{U}_2^*(s\mathbf{I} + \mathbf{P}_1)^{-1} \mathbf{U}_2)$$

where we have used the fact that

$$\det(s\mathbf{I} - \mathbf{A} - \mathbf{BC}) = \det(s\mathbf{I} - \mathbf{A}) \det(\mathbf{I} - \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1} \mathbf{B})$$

$$\eta(k) = \frac{\min_{m \neq n} \sqrt{\mathbf{h}_m^T(k) \mathbf{P}_{\mathbf{S}(k)}^\perp \mathbf{h}_m(k) [1 - \cos(\mathbf{h}_m(k), \mathbf{h}_n(k); \mathbf{P}_{\mathbf{S}(k)}^\perp)]}}{\min_{m \neq n} \sqrt{\mathbf{h}_m^T(k) \mathbf{h}_m(k) [1 - \cos(\mathbf{h}_m(k), \mathbf{h}_n(k))]} \quad (28)$$

$$\mathbf{P}_m = \begin{cases} \mathbf{h}_m(k)(\mathbf{h}_m^*(k)\mathbf{h}_m(k))^{-1}\mathbf{h}_m^*(k), & \text{SU case} \\ \mathbf{P}_{\mathbf{S}(k)}^\perp \mathbf{h}_m(k) \left( \mathbf{h}_m^*(k) \mathbf{P}_{\mathbf{S}(k)}^\perp \mathbf{h}_m(k) \right)^{-1} \mathbf{h}_m^*(k) \mathbf{P}_{\mathbf{S}(k)}^\perp, & \text{MU case.} \end{cases} \quad (29)$$

$$\Delta \mathbf{P} = \mathbf{P}_n - \mathbf{P}_m. \quad (30)$$

when all of the matrices are of appropriate size [14, Appendix A.13]. We can expand the first term as

$$\det(s\mathbf{I} + \mathbf{P}_1) = s^{N-r}(s+1)^r.$$

The second term may be expanded through the Woodbury identity

$$\det(\mathbf{I} - \mathbf{U}_2^*(s\mathbf{I} + \mathbf{U}_1\mathbf{U}_1^*)^{-1}\mathbf{U}_2) = s^{-r}(s+1)^{-r} \det((s-1)(s+1)\mathbf{I} + \mathbf{U}_2^*\mathbf{P}_1\mathbf{U}_2)$$

[14, Appendix A.20]. Collecting these terms we find

$$p(s) = s^{N-2r} \det(s\mathbf{I} + [\mathbf{I} - \mathbf{U}_2^*\mathbf{P}_1\mathbf{U}_2]^{1/2}) \cdot \det(s\mathbf{I} - [\mathbf{I} - \mathbf{U}_2^*\mathbf{P}_1\mathbf{U}_2]^{1/2}).$$

The eigenvalues of  $\mathbf{I} - \mathbf{U}_2^*\mathbf{P}_1\mathbf{U}_2$  are known to be  $\{\sin^2(\theta_i)\}$  [15] and the lemma follows.  $\square$

Factoring the projections into  $\mathbf{P}_m = \mathbf{U}_m\mathbf{U}_m^*$  and  $\mathbf{P}_n = \mathbf{U}_n\mathbf{U}_n^*$  it follows from Lemma V.1 that

$$\Delta \mathbf{P} = \mathbf{U}\Sigma\mathbf{U}^* = [\mathbf{u}_1, \mathbf{u}_2] \begin{bmatrix} s & 0 \\ 0 & -s \end{bmatrix} [\mathbf{u}_1, \mathbf{u}_2]^* \quad (31)$$

where  $s = |\sin(\theta)|$  and  $\theta$  is the angle between  $\mathbf{h}_m(k)$  and  $\mathbf{h}_n(k)$ . Using the definitions

$$\begin{aligned} z &= \mathbf{y}^* \Delta \mathbf{P} \mathbf{y} \\ \lambda_1 &= |\mu(k)|^2 \mathbf{h}_m^*(k) \mathbf{u}_1 \mathbf{u}_1^* \mathbf{h}_m(k) \\ \lambda_2 &= |\mu(k)|^2 \mathbf{h}_m^*(k) \mathbf{u}_2 \mathbf{u}_2^* \mathbf{h}_m(k) \end{aligned} \quad (32)$$

the following relationships hold:

$$\begin{aligned} \lambda_1 - \lambda_2 &= -\frac{|\mu(k)|^2}{s} \mathbf{h}_m^*(k) \mathbf{P}_n^\perp \mathbf{h}_m(k) \\ \lambda_1 + \lambda_2 &= |\mu(k)|^2 \mathbf{h}_m^*(k) \mathbf{h}_m(k) \\ 4\lambda_1\lambda_2 &= (|\mu(k)|^2 \mathbf{h}_m^*(k) \mathbf{h}_m(k))^2 \cos(\mathbf{h}_m(k), \mathbf{h}_n(k))^2. \end{aligned} \quad (33)$$

We find the characteristic function for the quadratic form  $z$  using the techniques of [16, Appendix B]:

$$G_z(\omega) = \frac{\exp\left\{-j\omega\left(\frac{\lambda_1}{1+j\omega s\sigma^2} - \frac{\lambda_2}{1-j\omega s\sigma^2}\right)\right\}}{(1+j\omega s\sigma^2)(1-j\omega s\sigma^2)}. \quad (34)$$

Following the method employed in [17, Appendix B] we find the probability of error to be

$$P[z > 0] = Q_1\left(\sqrt{\frac{\lambda_2}{2\sigma^2}}, \sqrt{\frac{\lambda_1}{2\sigma^2}}\right)$$

$$- \frac{1}{2} \exp\left\{-\frac{1}{4\sigma^2}(\lambda_1 + \lambda_2)\right\} I_0\left(\frac{\sqrt{\lambda_1\lambda_2}}{2\sigma^2}\right) \quad (35)$$

$$\begin{aligned} &= Q_1\left(\sqrt{\frac{|\mu(k)|^2 |\mathbf{u}_2^* \mathbf{h}_m(k)|}{2\sigma^2}}, \sqrt{\frac{|\mu(k)|^2 |\mathbf{u}_1^* \mathbf{h}_m(k)|}{2\sigma^2}}\right) \\ &\quad - \frac{1}{2} \exp\left(-\frac{|\mu(k)|^2 \|\mathbf{h}_m(k)\|^2}{4\sigma^2}\right) \\ &\quad \times I_0\left(\frac{|\mu(k)|^2 \|\mathbf{h}_m(k)\| \|\mathbf{h}_n(k)\|}{2\sigma^2 \|\mathbf{h}_n(k)\|}\right) \end{aligned} \quad (36)$$

where  $Q_1(a, b)$  is Marcum's  $Q$ -function and we have used the relationships listed in (33). We can obtain an asymptotically tight expression for (35) using the results of [18] to find

$$P[z > 0] \Rightarrow \sqrt{\frac{1+c}{4c}} Q\left(\sqrt{\frac{|\mu(k)|^2 \mathbf{h}_m^*(k) \mathbf{h}_m(k) (1-c)}{2\sigma^2}}\right) \quad (37)$$

as  $\sigma^2 \rightarrow 0$ , where  $c = +\sqrt{1-s^2}$ .

Building  $\underline{P}_{\text{SU}}(\sigma^2 r^{-1})$ ,  $\underline{P}_\phi(\sigma^2)$ ,  $\bar{P}_{\text{SU}}(\sigma^2 r^{-1})$ , and  $\bar{P}_\phi(\sigma^2)$  as in Section V-A we find the MAE in (38) at the bottom of this page. Hence the asymptotic efficiency is nonzero as long as the signal set for each user is not pairwise colinear in the interference-free subspace  $\langle \mathbf{S} \rangle$  (meaning  $|\cos(\mathbf{h}_m, \mathbf{h}_n; \mathbf{P}_{\mathbf{S}}^\perp)| < 1$ ), and not in the span of the interfering user's signals. Notice that the fading parameter,  $\mu(k)$ , does not affect the asymptotic efficiency. It does, however, effect the probability of error for a given value of  $\sigma^2$ . The upper and lower bounds would be obtained by averaging the expression of (37) over the distribution for  $|\mu(k)|$ .

## VI. MULTIUSER COMMUNICATION ON THE FREQUENCY SELECTIVE FADING CHANNEL

In the presence of frequency-selective fading with multipath spread  $T_{\text{mp}}$  (assumed to be equal for each user for simplicity) we develop a tapped-delay line model of the channel as in [17].

$$\eta(k) = \frac{\min_{m \neq n} \mathbf{h}_m^*(k) \mathbf{P}_{\mathbf{S}(k)}^\perp \mathbf{h}_m(k) \left(1 - \left|\cos(\mathbf{h}_m(k), \mathbf{h}_n(k); \mathbf{P}_{\mathbf{S}(k)}^\perp)\right|\right)}{\min_{m \neq n} \mathbf{h}_m^*(k) \mathbf{h}_m(k) (1 - |\cos(\mathbf{h}_m(k), \mathbf{h}_n(k))|)}. \quad (38)$$

The  $k$ th user transmits a signal  $\mathbf{s}(k) \in \mathfrak{S}(k) = \{\mathbf{s}_m(k)\}_{m=1}^M$ . Let  $W$  be the effective bandwidth of the  $\mathbf{s}_m(k)$ . The received signal has the complex baseband model

$$\mathbf{r} = \sum_{k=1}^K \sum_{j=1}^J c_{m,j}(k) \mathcal{S}^{j/W} \mathbf{s}_m(k) + \mathbf{n} \quad (39)$$

where the  $\{c_{j,m}(k)\}$  are the time-varying fading coefficients (assumed constant over one signal period) affecting the  $m$ th signal of the  $k$ th user,  $\mathcal{S}^\tau$  is the delay operator in  $\mathbb{L}^2$  with  $\{\mathcal{S}^\tau \mathbf{h}\}(t) = \mathbf{h}(t - \tau)$ ,  $\mathbf{n}$  is additive (complex) white noise, and  $J = \lceil T_{\text{imp}} W \rceil + 1$ . Let the set  $\{\mathcal{S}^{j/W} \mathbf{s}_m(k)\}_{m,k,j}$  have the orthonormal basis  $\{\mathbf{v}(n)\}_{n=1}^N$ , with  $N \leq JKM$ . Then the correlation receiver computes the vector statistic

$$\begin{aligned} \mathbf{y} &= \langle \mathbf{r}, \mathbf{V} \rangle \\ &= \sum_{k=1}^K \mathbf{H}(k) \mathbf{b}(k) + \mathbf{n} \\ &= \mathbf{H}(k) \mathbf{b}(k) + \mathbf{S}(k) \boldsymbol{\beta} + \mathbf{n}, \end{aligned}$$

where

$$\mathbf{H}(k) = [\mathbf{H}_1(k), \dots, \mathbf{H}_M(k)] \in \mathbb{C}^{N \times MJ}$$

and

$$\mathbf{H}_m(k) = [\langle \mathcal{S}^{1/W} \mathbf{s}_m(k), \mathbf{V} \rangle, \dots, \langle \mathcal{S}^{J/W} \mathbf{s}_m(k), \mathbf{V} \rangle] \in \mathbb{C}^{N \times J}.$$

The vector  $\mathbf{b}(k)$  contains just one nonzero  $J$ -vector

$$\mathbf{b}(k) \in \{\mathbf{E}_m \mathbf{c}_m(k)\}, \quad \mathbf{E}_m = [\mathbf{0} \dots \mathbf{I} \dots \mathbf{0}]^T \quad (40)$$

and  $\mathbf{c}_m(k)^T = [c_{1,m}(k), c_{2,m}(k), \dots, c_{J,m}(k)]$ .

We consider the case where the fading vectors  $\mathbf{c}_m(k)$  are varying sufficiently fast from symbol to symbol so as to make accurate estimation infeasible. This implies that the fading coherence time is commensurate with the signaling period. Then we might ask for decision rules which are invariant to the particular fading realization.

The GML detector for this problem is found as in the frequency-nonselctive case. First, the maximum-likelihood estimates of  $\mathbf{H}_m(k) \mathbf{c}(k)$  and  $\mathbf{S}(k) \boldsymbol{\beta}$  are found

$$\begin{aligned} \mathbf{H}_m(k) \widehat{\mathbf{c}}_m(k) &= \mathbf{E}_{\mathbf{H}_m(k) \mathbf{S}(k)} \mathbf{y} \\ \mathbf{S} \hat{\boldsymbol{\beta}}_m(k) &= \mathbf{E}_{\mathbf{S}(k) \mathbf{H}_m(k)} \mathbf{y}. \end{aligned} \quad (41)$$

These estimates are then substituted into the likelihood functions for the measurement  $\mathbf{y}$  and the corresponding GML rule is

$$\hat{m}(k) = \arg \max_m \mathbf{y}^* \mathbf{P}_{\mathbf{S}(k)}^\perp \mathbf{H}_m(k) \mathbf{y}. \quad (42)$$

This is a full generalization of the rank-1 detector of (22) to the case of multiple-rank signaling imposed by the frequency-selective fading (with the short coherence time assumption) and appears to be original. The GML detector is a *matched subspace detector* [11], [12] which seeks the subspace  $\langle \mathbf{P}_{\mathbf{S}(k)}^\perp \mathbf{H}_m(k) \rangle$  in which  $\mathbf{y}$  has the most energy. The detector is invariant to rotations and scaling within the subspace  $\langle \mathbf{H}_m(k) \rangle$  as well as to addition of vectors lying in  $\langle \mathbf{S}(k) \rangle$ . It should be noted that we require  $M$  linearly independent subspaces,  $\langle \mathbf{P}_{\mathbf{S}}^\perp \mathbf{H}_m(k) \rangle$ , in order to perform detection.

By letting  $\mathbf{G} = \mathbf{P}_{\mathbf{S}(k)}^\perp \mathbf{H}(k)$ , we can rewrite the quadratic form in (42) as

$$\mathbf{y}^* \mathbf{P}_{\mathbf{S}(k)}^\perp \mathbf{H}_m(k) \mathbf{y} = \mathbf{y}^* \mathbf{P}_{\mathbf{G}} \mathbf{H}_m(k) \mathbf{y}. \quad (43)$$

The detector may be written as

$$\hat{m}(k) = \arg \max_m \mathbf{y}^* \mathbf{P}_{\mathbf{G}} \mathbf{H}_m(k) \mathbf{y}. \quad (44)$$

The advantage of this form of the GML detector is that the projection matrices  $\mathbf{P}_{\mathbf{G}} \mathbf{H}_m(k)$  can be estimated with low computational complexity. This will be explored in Section VIII-C.

## VII. INTERFERENCE IDENTIFICATION

The detectors of (21), (22), and (42) require knowledge of the interference subspace  $\langle \mathbf{S}(k) \rangle$  for their implementations. This knowledge may be unavailable to the receiver or it may be computationally burdensome to track in a time-varying situation as users enter and exit the channel. To mitigate this problem, we seek low-complexity estimators of this subspace based on actual channel measurements. In the following section, we derive and analyze such an estimation procedure based on second-order statistics of the measurement. The results generalize a lemma of Wang and Poor [7] to NMM signaling and to OMM signaling over the frequency-selective fading channel.

### A. The Interference Identification Procedure

In this section we address the problem of identifying the interference subspace  $\langle \mathbf{S} \rangle$  from channel measurements. We assume that we have a set of measurements consisting of a stochastic signal constrained to lie in a *known* subspace  $\langle \mathbf{H} \rangle$ , plus a stochastic interference lying in an *unknown* subspace  $\langle \mathbf{S} \rangle$ , plus additive white noise. We assume that the signal and the interference terms are uncorrelated, as is the case in the multiple-access communication models that we have considered. We seek to identify  $\langle \mathbf{S} \rangle$  and will do so by deriving a technique for estimating  $\mathbf{P}_{\mathbf{S}}$ . Our result is a generalization of a theorem of Wang and Poor [7] with a proof derived from our representations of projections and pseudo-inverses in Section II.

*Theorem VII.1:* Given  $\mathbf{y} = \mathbf{H}\boldsymbol{\alpha} + \mathbf{S}\boldsymbol{\beta} + \mathbf{n}$  where  $\mathbf{H} \in \mathbb{C}^{N \times r}$  and  $\mathbf{S} \in \mathbb{C}^{N \times (p-r)}$  with  $\text{rank}([\mathbf{H}, \mathbf{S}]) = p$ . Let  $\mathbf{y}$  have the correlation matrix

$$\mathbf{R} = \mathbf{E}[\mathbf{y}\mathbf{y}^*] = \mathbf{H}\mathbf{R}_{\boldsymbol{\alpha}}\mathbf{H}^* + \mathbf{S}\mathbf{R}_{\boldsymbol{\beta}}\mathbf{S}^* + \sigma^2\mathbf{I}$$

where  $\mathbf{R}_{\boldsymbol{\alpha}} = \mathbf{E}[\boldsymbol{\alpha}\boldsymbol{\alpha}^*] > \mathbf{0}$ ,  $\mathbf{R}_{\boldsymbol{\beta}} = \mathbf{E}[\boldsymbol{\beta}\boldsymbol{\beta}^*] > \mathbf{0}$ , and  $\mathbf{E}[\mathbf{n}\mathbf{n}^*] = \sigma^2\mathbf{I}$ . The eigendecomposition of  $\mathbf{R}$  is

$$\mathbf{R} = [\mathbf{U}_{\mathbf{H}\mathbf{S}} \quad \mathbf{U}_A] \begin{bmatrix} \boldsymbol{\Lambda}_{\mathbf{H}\mathbf{S}}^2 + \sigma^2\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma^2\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{\mathbf{H}\mathbf{S}}^* \\ \mathbf{U}_A^* \end{bmatrix}$$

where

$$\mathbf{H}\mathbf{R}_{\boldsymbol{\alpha}}\mathbf{H}^* + \mathbf{S}\mathbf{R}_{\boldsymbol{\beta}}\mathbf{S}^* = \mathbf{U}_{\mathbf{H}\mathbf{S}} \boldsymbol{\Lambda}_{\mathbf{H}\mathbf{S}}^2 \mathbf{U}_{\mathbf{H}\mathbf{S}}^*.$$

Defining  $\mathbf{G} = \mathbf{U}_{\mathbf{H}\mathbf{S}} \boldsymbol{\Lambda}_{\mathbf{H}\mathbf{S}}^{-2} \mathbf{U}_{\mathbf{H}\mathbf{S}}^* \mathbf{H}$ , we have  $\langle \mathbf{G} \rangle = \langle \mathbf{P}_{\mathbf{S}}^\perp \mathbf{H} \rangle$  and  $\mathbf{P}_{\mathbf{G}} = \mathbf{P}_{\mathbf{P}_{\mathbf{S}}^\perp \mathbf{H}}$ .

*Proof:* See Appendix B.

From the identity  $\mathbf{P}_{\mathbf{H}\mathbf{S}} = \mathbf{P}_{\mathbf{S}} + \mathbf{P}_{\mathbf{P}_{\mathbf{S}}^\perp \mathbf{H}}$  it follows that

$$\begin{aligned} \mathbf{P}_{\mathbf{S}} &= \mathbf{P}_{\mathbf{H}\mathbf{S}} - \mathbf{P}_{\mathbf{P}_{\mathbf{S}}^\perp \mathbf{H}} \\ &= \mathbf{U}_{\mathbf{H}\mathbf{S}} \mathbf{U}_{\mathbf{H}\mathbf{S}}^* - \mathbf{P}_{\mathbf{G}}. \end{aligned} \quad (45)$$

Each of the terms in (45) may be adaptively estimated from measurements using an eigendecomposition of the sample covariance matrix for block processing or an iterative approximation to this decomposition using subspace tracking as described in Section VII-D.

Notice that we can drop the requirement that  $\mathbf{R}_{\alpha\alpha}$  and  $\mathbf{R}_{\beta\beta}$  be full rank in Theorem VII.1 by replacing  $\mathbf{H}\mathbf{R}_{\alpha\alpha}\mathbf{H}^*$  and  $\mathbf{S}\mathbf{R}_{\beta\beta}\mathbf{S}^*$  by their appropriate reduced dimensional representations. The results of the theorem would still hold.

### B. Perturbation Analysis

In this section we will find bounds for the distance between the estimated subspace  $\langle \hat{\mathbf{G}} \rangle$  and the true subspace  $\langle \mathbf{G} \rangle$ , when  $\hat{\mathbf{G}}$  is computed from a perturbed version of  $\mathbf{R}$ . We assume

$$\hat{\mathbf{R}} = \mathbf{R} + \mathbf{N} \quad (46)$$

where  $\mathbf{N}$  is a “small” matrix in the 2-norm sense.

Let  $\hat{\mathbf{R}}$  have the model of (46) and write the SVD of  $\hat{\mathbf{R}}$  as

$$\hat{\mathbf{R}} = [\hat{\mathbf{U}}_{\mathbf{H}\mathbf{S}} \quad \hat{\mathbf{U}}_{\mathbf{A}}] \begin{bmatrix} \hat{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \hat{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{U}}_{\mathbf{H}\mathbf{S}}^* \\ \hat{\mathbf{U}}_{\mathbf{A}}^* \end{bmatrix}. \quad (47)$$

The matrix  $\mathbf{G}$  and the noise power  $\sigma^2$  are estimated as

$$\begin{aligned} \hat{\mathbf{G}} &= \hat{\mathbf{U}}_{\mathbf{H}\mathbf{S}} (\hat{\Sigma}_1 - \hat{\sigma}^2 \mathbf{I})^{-1} \hat{\mathbf{U}}_{\mathbf{H}\mathbf{S}}^* \mathbf{H} \\ \hat{\sigma}^2 &= \frac{1}{N-p} \text{Tr}(\hat{\Sigma}_2) \end{aligned} \quad (48)$$

where  $p$  is the rank of the signal plus interference subspace. With this decomposition we find the following result.

**Lemma VII.2:** Define the matrices  $\hat{\mathbf{\Lambda}}_{\Delta} = \mathbf{\Lambda}_{\mathbf{H}\mathbf{S}}^2 - (\hat{\Sigma}_1 - \hat{\sigma}^2 \mathbf{I})$  and  $\hat{\mathbf{U}}_{\Delta} = \mathbf{U}_{\mathbf{H}\mathbf{S}} - \hat{\mathbf{U}}_{\mathbf{H}\mathbf{S}}$ . Let  $\|\mathbf{N}\|_2 = \delta$ . Then if the diagonal entries of  $\mathbf{\Lambda}_{\mathbf{H}\mathbf{S}}^2$ ,  $\{\lambda_i\}_{i=1}^p$ , are distinct<sup>2</sup> and  $\delta < \|\mathbf{\Lambda}_{\mathbf{H}\mathbf{S}}^2\|_2$ , and

$$\epsilon = \max \left\{ 2\delta, 4\sqrt{p} \max_{i \neq j} \left( \sqrt{1 + \left( \frac{\delta}{|\lambda_i - \lambda_j|} \right)^2} - 1 + \frac{\delta}{|\lambda_i - \lambda_j|} \right) \right\}$$

then  $\|\hat{\mathbf{\Lambda}}_{\Delta}\|_2 \leq \epsilon$  and  $\|\hat{\mathbf{U}}_{\Delta}\|_2 \leq \epsilon$ . The interpretation is that the error norms  $\|\hat{\mathbf{\Lambda}}_{\Delta}\|_2$  and  $\|\hat{\mathbf{U}}_{\Delta}\|_2$  are of the same order as  $\|\mathbf{N}\|_2$ .

*Proof:* See Appendix C.

We are interested in estimating  $\mathbf{G}$ , which requires an expression for  $(\hat{\Sigma}_1 - \hat{\sigma}^2 \mathbf{I})^{-1}$ . To this end, let the eigenvalues of  $\mathbf{\Lambda}_{\mathbf{H}\mathbf{S}}^2$  be  $\{\lambda_k\}$  and the diagonal entries of  $\hat{\Sigma}_1 - \hat{\sigma}^2 \mathbf{I}$  be given by  $\{\hat{\lambda}_k\}$ . Define the matrix  $\hat{\mathbf{\Lambda}}_1 = \text{diag}\{1/\hat{\lambda}_k - 1/\lambda_k\}$ , and note

<sup>2</sup>The restriction that the eigenvalues be distinct can be lifted. The dominant eigenvectors of  $\mathbf{R}$  will not be unique when there are multiple eigenvalues but any eigendecomposition of  $\hat{\mathbf{R}}$  will yield eigenvectors “close” to a valid decomposition of  $\mathbf{R}$ .

$(\hat{\Sigma}_1 - \hat{\sigma}^2 \mathbf{I})^{-1} = \mathbf{\Lambda}_{\mathbf{H}\mathbf{S}}^{-2} + \hat{\mathbf{\Lambda}}_1$ . The  $k$ th singular value of  $\hat{\mathbf{\Lambda}}_1$  is bounded by  $\epsilon/(\lambda_k - \epsilon)\lambda_k$  so let  $\bar{\lambda}$  maximize this bound over the set  $\{\lambda_k\}$ . Then  $\|\hat{\mathbf{\Lambda}}_1\|_2 \leq \epsilon/(\bar{\lambda}(\bar{\lambda} - \epsilon))$ . The estimate,  $\hat{\mathbf{G}}$ , of  $\mathbf{G}$  is given by

$$\hat{\mathbf{G}} = \hat{\mathbf{U}}_{\mathbf{H}\mathbf{S}} (\hat{\Sigma}_1 - \hat{\sigma}^2 \mathbf{I})^{-1} \hat{\mathbf{U}}_{\mathbf{H}\mathbf{S}}^* \mathbf{H} \quad (49)$$

$$= \mathbf{G} + \mathbf{N}_1 \quad (50)$$

where the matrix  $\mathbf{N}_1$  is found to be

$$\begin{aligned} \mathbf{N}_1 &= \mathbf{U}_{\mathbf{H}\mathbf{S}} \mathbf{\Lambda}_{\mathbf{H}\mathbf{S}}^{-2} \hat{\mathbf{U}}_{\Delta}^* \mathbf{H} + \mathbf{U}_{\mathbf{H}\mathbf{S}} \hat{\mathbf{\Lambda}}_1 \hat{\mathbf{U}}_{\mathbf{H}\mathbf{S}}^* \mathbf{H} \\ &\quad + \hat{\mathbf{U}}_{\Delta} (\mathbf{\Lambda}_{\mathbf{H}\mathbf{S}}^{-2} + \hat{\mathbf{\Lambda}}_1) \hat{\mathbf{U}}_{\mathbf{H}\mathbf{S}}^* \mathbf{H}. \end{aligned} \quad (51)$$

Since  $|\lambda_k - \hat{\lambda}_k| \leq \epsilon$  and  $\|\hat{\mathbf{U}}_{\Delta}\|_2 < \epsilon$  by construction we find that

$$\begin{aligned} \|\mathbf{N}_1\|_2 &\leq \|\mathbf{H}\|_2 \left( \epsilon \|\mathbf{\Lambda}_{\mathbf{H}\mathbf{S}}^{-2}\|_2 + \|\hat{\mathbf{\Lambda}}_1\|_2 + \epsilon \|\mathbf{\Lambda}_{\mathbf{H}\mathbf{S}}^{-2} + \hat{\mathbf{\Lambda}}_1\|_2 \right) \\ &\leq \epsilon \|\mathbf{H}\|_2 \left( \frac{1}{\bar{\lambda}} + \frac{1}{\bar{\lambda}(\bar{\lambda} - \epsilon)} + \frac{1}{\bar{\lambda} - \epsilon} \right) \end{aligned} \quad (52)$$

where  $\bar{\lambda} = \min_i \{\lambda_i\}$ .

Let the singular values of  $\mathbf{G}$  be  $\{\beta_k\}$ . Then we can bound the distance  $\text{dist}(\langle \mathbf{G} \rangle, \langle \hat{\mathbf{G}} \rangle)$  using the invariant subspace sensitivity results of [9, Sec. 7.2] to find

$$\text{dist}(\langle \mathbf{G} \rangle, \langle \hat{\mathbf{G}} \rangle) \leq 4 \frac{\|\mathbf{N}_1\|_2}{\min_k \{\beta_k\}}. \quad (53)$$

### C. Convergence

Having established bounds on  $\text{dist}(\langle \mathbf{G} \rangle, \langle \hat{\mathbf{G}} \rangle)$  through a perturbation analysis, we now examine the convergence of the subspace estimation algorithm. We will show that estimating  $\mathbf{G}$  based on an eigendecomposition of the sample correlation matrix from independent and identically distributed (i.i.d.) measurements with correlation matrix  $\mathbf{R}$  produces a strongly consistent estimate of the true matrix  $\mathbf{G}$  and we give asymptotic bounds on the distance measure.

We assume that we have  $Q$  independent measurements,  $\{\mathbf{y}_q\}_{q=1}^Q$ , each identically distributed with zero mean and correlation matrix

$$\mathbf{R} = \mathbf{H}\mathbf{R}_{\alpha\alpha}\mathbf{H}^* + \mathbf{S}\mathbf{R}_{\beta\beta}\mathbf{S}^* + \sigma^2 \mathbf{I}.$$

Compute the sample correlation matrix

$$\hat{\mathbf{R}} = \frac{1}{Q} \sum_{q=1}^Q \mathbf{y}_q \mathbf{y}_q^* \quad (54)$$

and eigendecompose  $\hat{\mathbf{R}}$  as in (47). The sample correlation matrix converges to  $\mathbf{R}$  almost surely (a.s.) by the Kolmogorov Law of Large Numbers [19] and hence [20]<sup>3</sup>

$$\hat{\mathbf{U}}_{\mathbf{H}\mathbf{S}} \rightarrow \mathbf{U}_{\mathbf{H}\mathbf{S}} \text{ a.s. and } \hat{\Sigma}_1 - \hat{\sigma}^2 \mathbf{I} \rightarrow \mathbf{\Lambda}_{\mathbf{H}\mathbf{S}}^2 \text{ a.s.} \quad (55)$$

Notice that the preceding equations are enough to show that  $\hat{\mathbf{G}} \rightarrow \mathbf{G}$  a.s. but that in general this is not sufficient to conclude that  $\text{dist}(\langle \mathbf{G} \rangle, \langle \hat{\mathbf{G}} \rangle) \rightarrow 0$ . Consider, for example, the non-stochastic sequence of matrices  $\mathbf{X}_k = \mathbf{X} + (1/2^k) \mathbf{v} \mathbf{v}^*$  for some

<sup>3</sup>In order to discuss the convergence of the eigenvector matrix,  $\hat{\mathbf{U}}_{\mathbf{H}\mathbf{S}}$ , we make the standard assumption that each eigenvector satisfies some uniqueness condition. We could, for example, constrain  $\{\mathbf{U}_{\mathbf{H}\mathbf{S}}\}_{n,n}$  to be real-valued for each  $n$ , see, e.g., [20].



TABLE I  
THE PASTd [27], [28] ALGORITHM FOR  
UPDATING THE DOMINANT CORRELATION STRUCTURE USING EXPONENTIAL  
AVERAGING,  $\hat{\mathbf{R}}(q) = \alpha \hat{\mathbf{R}}(q-1) + (1-\alpha)\mathbf{y}(q)\mathbf{y}^*(q)$ . THE "EXTRA"  
EIGENVALUE  $\lambda_{p+1}(q)$  IS USED TO UPDATE THE NOISE ESTIMATE,  $\hat{\sigma}^2$

<b>Inputs:</b>	$\mathbf{U}(q-1) \in \mathbb{C}^{N \times p}, \{\lambda_i(q-1)\}_{i=1}^p, \mathbf{y}(q), \alpha$
<b>Main Loop:</b>	
	$\mathbf{x}_1(q) = \mathbf{y}(q)$
FOR	$k = 1 : p$ DO
	$\beta_k(q) = \mathbf{u}_k^*(q-1)\mathbf{x}_k(q)$ $\lambda_k(q) = \alpha\lambda_k(q-1) +  \beta_k(q) ^2$ $\mathbf{u}_k(q) = \mathbf{u}_k(q-1) + \frac{\beta_k(1)\mathbf{x}_k(q) - \mathbf{u}_k(q-1)\beta_k(q)}{\lambda_k(q)}$ $\mathbf{x}_{k+1}(q) = \mathbf{x}_k(q) - \beta_k(q)\mathbf{u}_k(q)$
END	
	$\lambda_{p+1}(q) = \alpha\hat{\sigma}^2(q-1) + \ \mathbf{x}_{p+1}(q)\ ^2/(N-p)$ $\mathbf{u}_{p+1}(q) = \mathbf{x}_{p+1}(q)/\ \mathbf{x}_{p+1}(q)\ $
<b>Outputs:</b>	$\mathbf{U}(q) \in \mathbb{C}^{N \times (p+1)}$ $\{\lambda_i(q)\}_{i=1}^{p+1}$

low-rank matrix  $\mathbf{X}$  and vector  $\mathbf{v} \in \langle \mathbf{X} \rangle^\perp$ , then  $\mathbf{X}_k \rightarrow \mathbf{X}$  in the matrix norm but  $\langle \mathbf{X}_k \rangle = \langle [\mathbf{X}, \mathbf{v}] \rangle \not\rightarrow \langle \mathbf{X} \rangle$ . The perturbation bounds on this subspace distance (see (52) and (53)) developed in Section VII-B do allow us to prove such convergence as they bound  $\text{dist}(\langle \mathbf{G} \rangle, \langle \hat{\mathbf{G}} \rangle)$  asymptotically by a constant times the maximum of  $\|\hat{\mathbf{U}}_{\mathbf{H}\mathbf{S}} - \mathbf{U}_{\mathbf{H}\mathbf{S}}\|$  and  $\|\hat{\mathbf{A}}_{\mathbf{H}\mathbf{S}} - (\hat{\Sigma}_1 - \hat{\sigma}^2 \mathbf{I})\|$ . Each of these terms approaches zero almost surely and hence so does  $\text{dist}(\langle \mathbf{G} \rangle, \langle \hat{\mathbf{G}} \rangle)$ . The estimator is therefore strongly consistent.

It is known [21] that for i.i.d. samples the matrices  $\hat{\mathbf{A}}_\Delta$  and  $\hat{\mathbf{U}}_\Delta$  have asymptotic 2-norms

$$\begin{aligned} \|\hat{\mathbf{A}}_\Delta\|_2 &= O(\sqrt{\log \log Q/Q}) \quad \text{a.s.} \\ \|\hat{\mathbf{U}}_\Delta\|_2 &= O(\sqrt{\log \log Q/Q}) \quad \text{a.s..} \end{aligned}$$

We may use our bounds from Section VII-B together with these asymptotic expressions to find

$$\text{dist}(\langle \mathbf{G} \rangle, \langle \hat{\mathbf{G}} \rangle) = O(\sqrt{\log \log Q/Q}) \quad \text{a.s..} \quad (56)$$

#### D. Subspace Tracking for Low-Complexity Blind Detection

The subspace estimator of Theorem VII.1 requires estimates of the dominant eigenvectors  $\mathbf{U}_{\mathbf{H}\mathbf{S}}$  and dominant eigenvalues  $\{\lambda_i\}$  of the measurement correlation matrix  $\mathbf{R}$  and of the additive noise power  $\sigma^2$ . When the channel is time-variant, for example, when users are entering and departing a multiple-access system and their average received powers are varying due to motion of the transmitters relative to the receiver, we might estimate the measurement correlation either from a block of data or with a running window estimator of the form

$$\hat{\mathbf{R}}(q) = \alpha \hat{\mathbf{R}}(q-1) + (1-\alpha)\mathbf{y}(q)\mathbf{y}^*(q).$$

In either case, an eigendecomposition of  $\hat{\mathbf{R}}$  must be performed. There are several computationally efficient methods for updating the estimates of the parameters of interest ( $\mathbf{U}_{\mathbf{H}\mathbf{S}}, \mathbf{A}_{\mathbf{H}\mathbf{S}}, \sigma^2$ ) through subspace tracking, see, e.g., [22]–[28]. A key result from this literature is that it is possible

to track these parameters with a complexity of  $O(Np)$ , where  $N$  is the dimensionality of  $\mathbf{R}$ , and  $p$  is the dimension of the dominant subspace.

For the numerical results presented in Section IX we use the Projection Approximation Subspace Tracking with Deflation (PASTd) algorithm [27], [28] to perform the subspace tracking. The PASTd algorithm is an  $O(Np)$  algorithm and is summarized in Table I.

Notice that this algorithm increases the rank of the dominant subspace by 1 at each iteration. If the true rank is known we allow the PASTd algorithm to perform this increase until  $p$  equals the true rank and then prune off the extra eigenvalue and eigenvector after each iteration. The noise estimate is updated as

$$\hat{\sigma}^2(q) \leftarrow \frac{1}{N-p}((N-p-1)\alpha\hat{\sigma}^2(q-1) + \lambda_{p+1}(q)). \quad (57)$$

When the true rank of the dominant subspace is not known we employ the PASTd algorithm together with a rank estimation procedure such as the Akaike Information Criterion (AIC) [29], which chooses the rank by

$$\hat{p} = \arg \max_k \left\{ \frac{N-k}{1-\alpha} \log \left( \frac{1/(N-k) \sum_{i=k+1}^N \lambda_i(q)}{\left( \prod_{i=k+1}^N \lambda_i(q) \right)^{1/(N-k)}} \right) + k(2N-k) \right\} \quad (58)$$

where  $1/(1-\alpha)$  is the effective number of measurements.

There are three possible actions based on the outcome of this test.

- 1) ( $\hat{p} > p$ ) we allow the rank to increase by keeping  $\{\lambda_i(q)\}_1^{p+1}$  and  $\mathbf{U}(q)$  and update the noise power with

$$\hat{\sigma}^2(q) \leftarrow \alpha\hat{\sigma}^2(q-1);$$

- 2) ( $\hat{p} = p$ ) we keep  $\{\lambda_i(q)\}_1^p$  and the first  $p$  columns of  $\mathbf{U}(q)$  and update the noise power with (57);

- 3) ( $\hat{p} < p$ ) we keep  $\{\lambda_i(q)\}_1^{p-1}$  and the first  $p-1$  columns of  $\mathbf{U}(q)$  and update the noise power with

$$\begin{aligned} \hat{\sigma}^2(q) \leftarrow \frac{1}{N-p+1}((N-p-1)\alpha\hat{\sigma}^2(q-1) \\ + \lambda_p(q) + \lambda_{p+1}(q)). \end{aligned}$$

## VIII. BLIND MULTIUSER COMMUNICATIONS

We now use these interference subspace identification procedures to find *blind* implementations of the decentralized detectors developed in Sections III and VI. This extends the blind decorrelating detector of Wang and Poor [7] to multidimensional signal sets arising from NMM and frequency-selective fading. We show that the computational complexity of the blind detectors is of the same order as that of the subspace tracking problem, which can be solved efficiently. This shows that the blind detectors are feasible even at high data rates.

### A. Blind Detection for Frequency Nonselective Fading

When we do not know the interference subspace  $\langle \mathbf{S}(k) \rangle$ , we can use the results of Section VII to estimate it. We require a few additional assumptions in order to make the correlation matrix for  $\mathbf{y}$  meet the conditions of Theorem VII.1. We list the additional assumptions for the coherent and noncoherent cases separately.

When the channel is coherent we assume that the signal set for each user is symmetric about  $\mathbf{0}$  in the mean, e.g.,  $\mathbf{E}[\mathbf{H}(k)\mathbf{b}] = \mathbf{0}$ . This forces the cross correlation to be  $\mathbf{0}$

$$\begin{aligned} \mathbf{E}[\mu(k)\mathbf{H}(k)\mathbf{b}(k)\mathbf{b}(l)^T\mathbf{H}(l)^T\mu(l)] \\ = \mathbf{E}[\mu(k)\mathbf{H}(k)\mathbf{b}(k)]\mathbf{E}[\mu(l)\mathbf{b}(l)^T\mathbf{H}(l)^T] \\ \text{(the users signal independently)} \\ = \mathbf{0}. \end{aligned}$$

When the channel is noncoherent we need to be a little more careful. Proceeding as in the coherent case we require

$$\mathbf{E} \left[ |\mu(k)| \sum_{m=1}^M e^{j\phi_m(k)} \mathbf{h}_m(k) \right] = \mathbf{0}.$$

In order to satisfy this condition we will assume that the phase angle  $\phi_m(k)$  is uniformly distributed on  $[0, 2\pi]$  for each  $m$ . To reiterate, the noncoherent detector given in (21) was derived without any assumptions on the complex gain but the blind detector requires the assumption that  $\mu_m(k)$  is a zero-mean complex random variable. This assumption is met for Rayleigh fading channels, for example.

With these assumptions we find the correlation matrix for  $\mathbf{y}$  to be

$$\begin{aligned} \mathbf{R} &= \mathbf{E}[\mathbf{y}\mathbf{y}^*] \\ &= \mathbf{E}[|\mu(k)|^2]\mathbf{H}(k)\mathbf{H}(k)^* + \mathbf{S}(k)\mathbf{R}_{\beta\beta}\mathbf{S}(k)^* + \sigma^2\mathbf{I}. \end{aligned} \quad (59)$$

If  $\mu(k)$  is modeled as an unknown deterministic quantity for the coherent channel, then  $\mathbf{E}[|\mu(k)|^2] = |\mu(k)|^2$ .

It may be the case that the matrix  $\mathbf{H}(k)$  is rank deficient (for example, when phase-shift keying is employed). Since we know  $\mathbf{H}(k)$  we may change the model in (59) to

$$\mathbf{R} = \mathbf{E}[|\mu(k)|^2]\tilde{\mathbf{H}}(k)\tilde{\mathbf{H}}(k)^* + \mathbf{S}(k)\mathbf{R}_{\beta\beta}\mathbf{S}(k)^* + \sigma^2\mathbf{I} \quad (60)$$

where  $\tilde{\mathbf{H}}(k)$  has full column rank.

We see that our correlation matrix satisfies the conditions of Theorem VII.1. This means that we may collect a time series of measurements  $\{\mathbf{y}(q)\}$  and use them to form the sample correlation matrix  $\hat{\mathbf{R}}$  and its eigendecomposition. These terms are then used to form the estimate  $\hat{\mathbf{G}}$  and the projection  $\mathbf{P}_{\hat{\mathbf{G}}}$  is used in the detector of (22).

### B. Blind Detection for Frequency Selective Fading

For frequency selective fading recall that our model for the measurement is

$$\mathbf{y} = \mathbf{H}(k)\mathbf{b}(k) + \mathbf{S}(k)\boldsymbol{\beta} + \mathbf{n}.$$

Let  $\mathbf{S}(k)\boldsymbol{\beta}$  have the correlation matrix  $\mathbf{S}(k)\mathbf{R}_{\beta\beta}\mathbf{S}(k)^*$  and let  $\mathbf{c}(k)$  be distributed complex normal  $\mathbf{c}(k) \sim \mathcal{CN}[\mathbf{0}, \mathbf{R}_{cc}(k)]$ . Then  $\mathbf{y}$  has the correlation

$$\mathbf{R} = \mathbf{H}(k)\mathbf{R}_{bb}(k)\mathbf{H}(k)^* + \mathbf{S}(k)\mathbf{R}_{\beta\beta}\mathbf{S}(k)^* + \sigma^2\mathbf{I} \quad (61)$$

where, from (40),

$$\mathbf{R}_{bb}(k) = 1/M \text{diag}\{\mathbf{R}_{cc}(k), \dots, \mathbf{R}_{cc}(k)\}$$

is a block-diagonal matrix with  $M$  blocks. We see that our data fits the conditions of Theorem VII.1 and we can use measurements of  $\mathbf{y}$  to estimate  $\mathbf{P}_{\mathbf{G}}$  and hence blindly build the detector of (44).

### C. Computational Complexity

In order for a detection scheme to be feasible for a communications channel it must be possible to implement it with low computational complexity, corresponding to high data rates. In this section we examine the computational complexity of the blind schemes developed in the previous two sections and show that when the correlation matrix for all users has rank  $r$  and the measurement size is  $N$ , the procedure has complexity  $O(Nr)$ .

First consider the detectors for frequency-nonselective fading in (21) and (22). Notice that for each  $m$

$$\begin{aligned} \mathbf{P}_{\mathbf{S}(k)}^\perp \mathbf{h}_m(k) &= \mathbf{P}_{\mathbf{P}_{\mathbf{S}(k)}^\perp \mathbf{H}(k)} \mathbf{h}_m(k) \\ &= \mathbf{U}_{\mathbf{G}} \mathbf{U}_{\mathbf{G}}^* \mathbf{h}_m(k) \end{aligned} \quad (62)$$

where  $\mathbf{U}_{\mathbf{G}}$  is defined in Section VII. Consequently, we need only find  $\mathbf{U}_{\hat{\mathbf{G}}}$ , the left singular vectors of  $\hat{\mathbf{G}}$ .

When subspace tracking is employed to iteratively decompose  $\mathbf{R}$  we may update the subspace matrices  $\mathbf{U}_{\mathbf{HS}}$  and  $\mathbf{\Lambda}_{\mathbf{HS}}$  with computational cost  $O(Nr)$  using an efficient update algorithm of the type described in Section VII-D. The matrix multiply to find  $\hat{\mathbf{G}}$  has complexity  $O(Nr)$ . The decomposition of  $\hat{\mathbf{G}}$  to get an orthonormal basis for  $\langle \hat{\mathbf{G}} \rangle$  has complexity  $O(N\tilde{M}^2)$  when an efficient technique is employed as described in [9, Ch. 5], where  $\mathbf{H}$  has rank  $\tilde{M}$ . The overall computational complexity is hence  $O(NP)$  where  $P = \max(\tilde{M}^2, r)$ .

The computational complexity of the blind detector for frequency-selective fading is found as in the frequency-nonselective fading case. The matrix  $\mathbf{U}_{\mathbf{G}}$  may be estimated with complexity  $O(N(MJ+L))$ , where  $L = \text{rank}(\mathbf{S})$  and the right singular vectors of  $\mathbf{U}_{\mathbf{G}}^* \mathbf{H}_m(k)$  may be estimated with complexity  $O(J^3)$ . When  $J^3$  is small relative to  $N(MJ+L)$  we find the overall complexity to be  $O(N(MJ+L))$ .

## IX. SIMULATION EXAMPLE

In order to gauge the effectiveness of the blind detector we examine the residual interference energy defined by  $\text{Tr}[\mathbf{U}_{\hat{\mathbf{G}}}^* \mathbf{S} \mathbf{R}_{\beta\beta} \mathbf{S}^* \mathbf{U}_{\hat{\mathbf{G}}}]$  for the noncoherent detector with Rayleigh fading. The user gains  $\{\mu_m(k)\}$  were taken to be complex normal random variables with variance  $\sigma_k^2 = \mathbf{E}[\mu(k)|^2]$ . The signal-to-noise ratio (SNR) of user 1 was 26 dB. Each user employed  $M = 3$  with length  $N = 31$  Gold codes (user 1 used codes 5–7, user 2 used codes 8–10, etc.) normalized to have unit norm.

At iteration 1 there were two interfering users each with energy levels  $(\mathbf{E}[\mu(l)|^2]/\mathbf{E}[\mu(1)|^2])$  of 14 dB relative to user 1. Fig. 3 shows the residual interference energy as a function of iteration number, averaged over 50 simulations, employing the

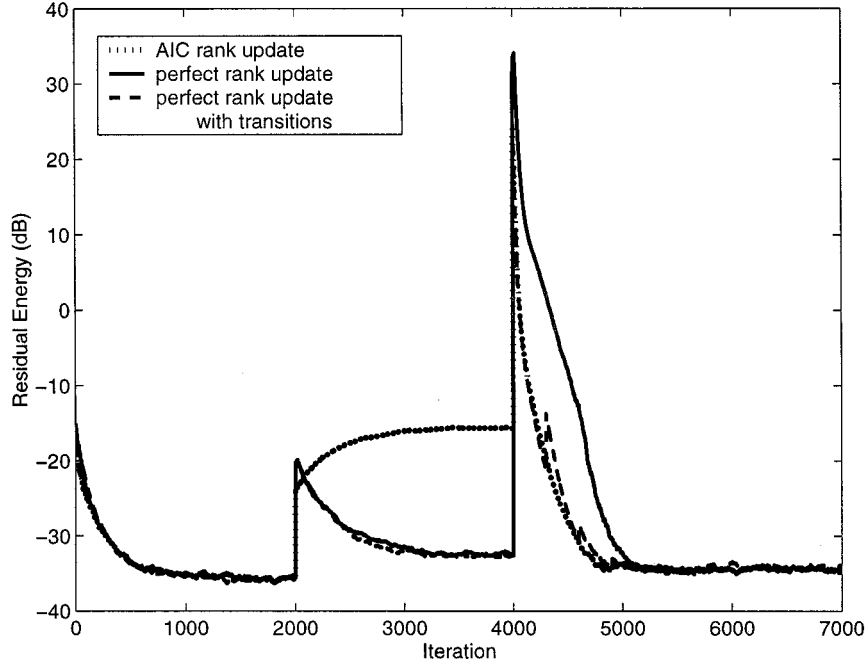


Fig. 3. Residual energy versus iteration number for the example of Section IX. At time  $k = 0$  there are two interfering users, each with energy levels of 14 dB relative to user 1. At time  $k = 2000$ , a third user with energy level  $-6$  dB enters the channel. At time  $k = 4000$ , the  $-6$ -dB and one of the 14-dB users exit the channel and a new user with energy level 44 dB enters.

PASTd subspace tracking algorithm together with the interference estimation. At the 2000th iteration, a third interfering user was added to the channel with an energy level of  $-6$  dB relative to user 1. At iteration number 4000 the  $-6$ -dB user and one of the 14-dB users left the channel and a 44-dB user entered. The simulations were run with the receiver in three different modes of operation; 1) automatic rank adjustment using the AIC, 2) perfect rank information, and 3) perfect rank information with a transition period following each rank adjustment. During this transition period the rank is held fixed. The third mode of operation allows a smooth transition when interfering users enter or exit the channel. The advantage of this is seen in the figure.

Notice that the rank-adaptive mode allows for good transient response but poor steady-state estimation. This is especially evident when the  $-6$ -dB user enters the channel. This user contributes little to the eigenvalues of the estimated correlation matrix and the AIC underestimates the rank of  $\hat{\mathbf{R}}$ , causing a constant amount of user power to remain unaccounted for.

The receiver with perfect rank information has a good steady state but poor transient response. This is especially evident when the 44-dB user enters the channel. In effect, the estimator forces the rank to be small when the estimated covariance matrix is far from its mean; this forces a large mismatch between the estimated and the true eigenstructure, allowing for power to be distributed in the subdominant eigenstructure. In steady state, the estimated covariance matrix has an eigenstructure much closer to its mean and the powerful user is then well accounted for by  $\hat{\mathbf{U}}_{\mathbf{H}\mathbf{S}}$ .

The third mode of operation attempts to improve the transient dynamics of the perfect rank estimator by allowing each rank reduction to occur after a set transition period (in this example 300 iterations were used). Increases in rank are allowed to

occur without this restriction since they do not suffer from the eigenstructure mismatch of the rank reductions. This controls the dominant eigenstructure mismatch observed in the second mode of operation.

## X. THE MMSE DETECTOR

In this paper we have concentrated on interference-canceling (or zero-forcing) detection of a desired user in multiple-access interference, wherein the interfering users' signals are completely nulled out by the interference nulling transformation  $\mathbf{P}_S^\perp$  (or  $\mathbf{P}_G$ ) prior to detection. We may also consider the minimum mean-square error (MMSE) detector in these scenarios in a manner analogous to the linear modulation case of [30].

For the case of flat fading, we return to the measurement model

$$\mathbf{y} = \mathcal{H}\mathcal{D}\mathbf{b} + \mathbf{n} \quad (63)$$

where  $\mathcal{H}$ ,  $\mathcal{D}$ , and  $\mathbf{b}$  are defined for the coherent and noncoherent channels in Sections III-A and III-B, respectively. We seek the MMSE estimate of  $\mathcal{D}\mathbf{b}$  and make a decision on user  $k$  by examining the  $k$ th block of the estimate  $\hat{\mathcal{D}}\mathbf{b}$ .

It is straightforward to show that the MMSE estimator for the coherent model is

$$\begin{aligned} \hat{\mathcal{D}}\mathbf{b} = & \mathcal{D} \left( \mathbf{R}_{bb} - \frac{1}{M^2} \mathbf{1}\mathbf{1}^T \right) \mathcal{D}^T \mathcal{H}^T \\ & \left( \sigma^2 \mathbf{I} + \mathcal{H} \mathcal{D} \left( \mathbf{R}_{bb} - \frac{1}{M^2} \mathbf{1}\mathbf{1}^T \right) \mathcal{D}^T \mathcal{H}^T \right)^{-1} \\ & \left( \mathbf{y} - \frac{1}{M} \mathcal{H} \mathcal{D} \mathbf{1} \right) + \frac{1}{M} \mathcal{D} \mathbf{1} \end{aligned} \quad (64)$$

where  $\mathbf{R}_{\mathbf{b}\mathbf{b}}$  is defined in Section III-A. Denote the  $k$ th length  $M$  subvector of  $\widehat{\mathcal{D}\mathbf{b}}$  by  $\widehat{\mathcal{D}\mathbf{b}}(k)$ . The MMSE detector for user  $k$  makes the decisions

$$\hat{m}_{\text{MMSE}}^C = \arg \max_m \{\widehat{\mathcal{D}\mathbf{b}}(k)\}_m. \quad (65)$$

When the users' signals satisfy the symmetry property

$$\sum \mathbf{h}_m(k) = \mathbf{H}(k) \mathbf{1} = \mathbf{0}$$

then we have  $\mathbf{R}_{\mathbf{b}\mathbf{b}} \mathcal{D}^T \mathcal{H}^T = (1/M) \mathcal{D}^T \mathcal{H}^T$  and  $\mathcal{H} \mathcal{D} \mathbf{1} = \mathbf{0}$ . In this case, we may simplify the MMSE detector to the matched filter rule

$$\hat{m}_{\text{MMSE}}^C(k) = \arg \max_m \{\mathbf{h}_m^T(k) \mathbf{R}^{-1} \mathbf{y}\} \quad (66)$$

where  $\mathbf{R} = \mathbf{E}[\mathbf{y}\mathbf{y}^*]$  is the measurement correlation matrix.

For the case of noncoherent signaling we assume as before that the fading terms  $\mu^{(m)}(k)$  are zero-mean complex random variables. The MMSE estimator is

$$\widehat{\mathcal{D}\mathbf{b}} = \mathbf{F} \mathbf{H}^* (\sigma^2 \mathbf{I} + \mathbf{H}^* \mathbf{F} \mathbf{H})^{-1} \mathbf{y} \quad (67)$$

where  $\mathbf{F}$  is defined in Section III-B. This leads to the simple decision rule for the noncoherent channel

$$\hat{m}_{\text{MMSE}}^{\text{NC}}(k) = \arg \max_m \{|\mathbf{h}_m^*(k) \mathbf{R}^{-1} \mathbf{y}|^2\}, \quad (68)$$

where  $\mathbf{R} = \mathbf{E}[\mathbf{y}\mathbf{y}^*]$ . Geometrically, we see that the noncoherent detector seeks the *whitened* signal vector  $\mathbf{R}^{-1/2} \mathbf{h}_m(k)$  which is closest to the *whitened* measurement  $\mathbf{R}^{-1/2} \mathbf{y}$  in terms of the magnitude squared inner product.

Notice that the MMSE detectors require knowledge of the fading parameters  $\mu(k)$ , or their correlation structure if they are stochastic and hence are *not* invariant to unknown measurement gain. It is interesting to note that asymptotically (as the background noise power goes to zero), the term  $\mathbf{H}(k)^* \mathbf{R}^{-1} \mathbf{y}$  approaches  $(\mathbf{H}(k)^* \mathbf{P}_{\mathcal{S}}^\perp \mathbf{H}(k))^{-1} \mathbf{H}(k)^* \mathbf{P}_{\mathcal{S}}^\perp \mathbf{y}$ , as can be verified through the Woodbury identity [14, A.21]. This means that the MMSE detectors do *not* approach the GML rules of (21) and (22), except for the case of diagonal  $\mathbf{H}(k)^* \mathbf{P}_{\mathcal{S}}^\perp \mathbf{H}(k)$ . In the rank-one case (linear modulation), we have  $\mathbf{H}(k) = \mathbf{h}(k)$  and the MMSE detector will always approach the GML solution (as is well known, see e.g. [30]). The conclusion is that the two techniques can yield different asymptotic performances (and hence asymptotic efficiencies) for the case of NMM. This problem is studied in [31] and the two techniques are shown to have different asymptotic performance characteristics, with the MMSE outperforming the GML detector for binary modulation (this does not generalize to larger cardinality constellations). In the following section we present a numerical example in which the GML detector outperforms the MMSE detector asymptotically for Rayleigh fading.

The blind implementation of the MMSE detectors is straightforward, requiring only the tracking of  $\hat{\mathbf{R}}^{-1}$ , which can be efficiently updated using the rank-one update of a matrix inverse (the Woodbury identity, [14, A.21])

$$\begin{aligned} & \hat{\mathbf{R}}^{-1}(q+1) \\ &= \frac{1}{\alpha} \left( \hat{\mathbf{R}}^{-1}(q) - \frac{\hat{\mathbf{R}}^{-1}(q) \mathbf{y}(q+1) \mathbf{y}^*(q+1) \hat{\mathbf{R}}^{-1}(q)}{\frac{\alpha}{1-\alpha} + \mathbf{y}^*(q+1) \hat{\mathbf{R}}^{-1}(q) \mathbf{y}(q+1)} \right) \end{aligned} \quad (69)$$

when  $\hat{\mathbf{R}}$  is updated through the exponential window

$$\hat{\mathbf{R}}(q+1) = \alpha \hat{\mathbf{R}}(q) + (1-\alpha) \mathbf{y}(q+1) \mathbf{y}^*(q+1).$$

Alternatively, we can employ a subspace-tracking procedure as in the blind GML detectors since

$$\mathbf{H}^* \mathbf{R}^{-1} = \mathbf{H}^* \mathbf{U}_{\mathbf{H}\mathbf{S}} (\Lambda_{\mathbf{H}\mathbf{S}}^2 + \sigma^2 \mathbf{I})^{-1} \mathbf{U}_{\mathbf{H}\mathbf{S}}^*.$$

This is the  $M$ -ary extension of the blind MMSE algorithm derived in [7].

We end this section by noting that the MMSE detector for the frequency-selective channel of Section VI can be derived in the same manner as the noncoherent detector for flat fading. In this case, we require knowledge of the joint correlation structure of the fading parameters  $\mathbf{E}[\mathbf{c}_m(k) \mathbf{c}_n^*(k)]$  for each user.

#### A. Simulation Example Comparing the MMSE and the GML Detectors

Using the same signal sets as in the example of Section IX we consider  $K = 3$  users with  $M = 3$  Gold sequences of length  $N = 31$  for each user. The interfering users each have a power of 14 dB relative to the user of interest. In Fig. 4 we plot the probability of error for the MMSE detector of (68) and the GML detector of (22). Each point was generated with 50 000 measurement realizations. Also plotted is the union upper bound on the probability of error for the GML detector. This bound is found by first obtaining the union upper bound conditioned on a particular  $|\mu(k)|$  as

$$P^U(|\mu(k)|) = \frac{1}{M} \sum_{m=1}^M \sum_{\substack{n=1 \\ n \neq m}}^M P[\mathbf{y}^* \Delta \mathbf{P} \mathbf{y} < 0 | \text{under hypothesis } m] \quad (70)$$

where  $\Delta \mathbf{P}$  is defined in (30) and the pairwise error probabilities are given by (37). This expression is then averaged against the Rayleigh distribution of  $|\mu(k)|$  to find the upper bound

$$P^U = \int_0^\infty \frac{2|\mu(k)|}{\mathbf{E}|\mu(k)|^2} e^{-|\mu(k)|^2 / \mathbf{E}|\mu(k)|^2} P^U(|\mu(k)|) d|\mu(k)|. \quad (71)$$

This bound was evaluated numerically.

Also shown in the figure is the probability of error when blind detection is performed. The blind detectors were built as described in Sections VIII and X from the eigendecomposition of the sample correlation matrix built from a block of 200 channel measurements. Each point on the blind curves is the average of 50 independent Monte Carlo simulations of the detectors (with different realizations of the sample-correlation matrix). Notice that the GML and the MMSE detector have almost identical performance for this channel. This is also true for the blind detectors. Notice further that the performance of both blind detectors approach a noise floor induced by the finite sample estimation of the correlation matrix.

#### XI. CONCLUSION

The contributions of this paper are threefold. First, we reviewed the theory of generalized maximum-likelihood detectors for multiple-access communication with nonlinear signaling

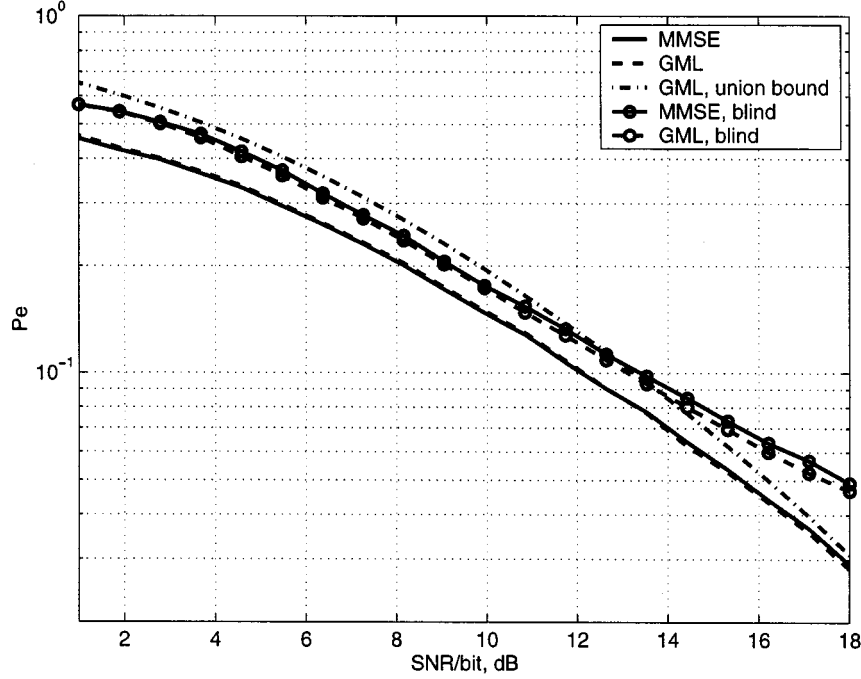


Fig. 4. Probability of a symbol error versus the output SNR per bit for the Rayleigh fading example of Section X-A. Each point is generated from 50 000 measurement realizations.

for both the coherent and the noncoherent frequency-nonselective fading channel and for the frequency-selective fading channel with a short coherence time. These detectors were posed in their most general form (without any requirements on the invertibility of the joint user signal correlation matrix  $\mathbf{E}[\mathbf{H}(1) \cdots \mathbf{H}(K)][\mathbf{H}(1) \cdots \mathbf{H}(K)]^*$ ). Geometrical interpretations of these detectors were presented and the relevant restrictions on the signal sets were derived.

We next developed a technique for determining an unknown interference subspace from channel measurements in a multidimensional signal plus multidimensional interference plus additive noise problem. This is useful not only for multiple-access communication, but for many statistical signal processing problems, including sensor array processing and Gauss–Gauss detection. The estimate is shown to be strongly consistent through a perturbation analysis and an asymptotic expression for the subspace fitting error was derived.

We applied these results to find blind detectors for multiuser, fading communication channels. The detectors are computationally efficient and hence feasible for high-speed data communication. These detectors build upon and generalize the recent blind decorrelating detector of Wang and Poor [7] for linear modulation.

Finally, the MMSE detectors for NMM signaling were presented. The MMSE detectors require knowledge of either the channel gains (for the coherent channel) or their statistics (for the noncoherent channel), which place them in a different category from the GML detectors which were designed to be invariant to these gains. In the case of the noncoherent detector, however, it was noted that a blind implementation of the MMSE detector is possible which does not require such knowledge. It was also shown that the GML is *not* generally the high SNR

limit of the MMSE detector for NMM signaling, an issue which is further explored in [31].

#### APPENDIX A PROOF OF THEOREM II.1

Following a method similar to that used in [8] we may first decompose  $\mathbf{P}_S$  as

$$\begin{aligned} \mathbf{P}_S &= [\mathbf{s}_1 \quad \mathbf{s}_2 \quad \cdots \quad \mathbf{s}_M](\mathbf{S}^* \mathbf{S})^{-1} \mathbf{S}^* \\ &= [\mathbf{s}_1 \quad \mathbf{0}](\mathbf{S}^* \mathbf{S})^{-1} \mathbf{S}^* + [\mathbf{0} \quad \mathbf{s}_2 \quad \mathbf{0}](\mathbf{S}^* \mathbf{S})^{-1} \mathbf{S}^* + \cdots \\ &\quad + [\mathbf{0} \quad \mathbf{s}_k \quad \mathbf{0}](\mathbf{S}^* \mathbf{S})^{-1} \mathbf{S}^* + \cdots \\ &\quad + [\mathbf{0} \quad \mathbf{s}_M](\mathbf{S}^* \mathbf{S})^{-1} \mathbf{S}^*. \end{aligned}$$

Let us now write  $\mathbf{S}$  as  $\mathbf{S} = [\mathbf{S}_{k,1} \quad \mathbf{s}_k \quad \mathbf{S}_{k,2}]$ , where  $\mathbf{S}_{k,1} = [\mathbf{s}_1 \quad \mathbf{s}_2 \quad \cdots \quad \mathbf{s}_{k-1}]$  and  $\mathbf{S}_{k,2} = [\mathbf{s}_{k+1} \quad \cdots \quad \mathbf{s}_M]$ . Then expand the  $k$ th term of  $\mathbf{P}_S$  as follows:

$$\begin{aligned} &[\mathbf{0} \quad \mathbf{s}_k \quad \mathbf{0}](\mathbf{S}^* \mathbf{S})^{-1} \mathbf{S}^* \\ &= [\mathbf{0} \quad \mathbf{s}_k \quad \mathbf{0}] \left( \begin{bmatrix} \mathbf{S}_{k,1}^* \mathbf{S}_{k,1} & \mathbf{S}_{k,1}^* \mathbf{s}_k & \mathbf{S}_{k,1}^* \mathbf{S}_{k,2} \\ \mathbf{s}_k^* \mathbf{S}_{k,1} & \mathbf{s}_k^* \mathbf{s}_k & \mathbf{s}_k^* \mathbf{S}_{k,2} \\ \mathbf{S}_{k,2}^* \mathbf{S}_{k,1} & \mathbf{S}_{k,2}^* \mathbf{s}_k & \mathbf{S}_{k,2}^* \mathbf{S}_{k,2} \end{bmatrix} \right)^{-1} \\ &\quad \times \begin{bmatrix} \mathbf{S}_{k,1}^* \\ \mathbf{s}_k^* \\ \mathbf{S}_{k,2}^* \end{bmatrix} \\ &= (\mathbf{s}_k^* \mathbf{s}_k)^{-1} [\mathbf{s}_k \quad \mathbf{0}] \\ &\quad \times \left[ \frac{1 + \frac{1}{(\mathbf{s}_k^* \mathbf{s}_k)} \mathbf{s}_k^* (\mathbf{S}_{k,1}^* \mathbf{P}_{\mathbf{s}_k}^\perp \mathbf{S}_{k,1})^{-1} \mathbf{S}_{k,1}^*}{\mathbf{X}} \mid \frac{\mathbf{s}_k^* (\mathbf{S}_{k,2}^* \mathbf{P}_{\mathbf{s}_k}^\perp \mathbf{S}_{k,2})^{-1}}{\mathbf{Z}} \right] \\ &\quad \times \begin{bmatrix} \mathbf{s}_k^* \\ \mathbf{S}_{k,2}^* \end{bmatrix} \end{aligned}$$

(permuting each term and applying the Matrix Inversion Lemma [12])

$$\begin{aligned} &= \mathbf{P}_{\mathbf{s}_k} \left( \mathbf{I} - \mathbf{S}_k \left( \mathbf{S}_k^* \mathbf{P}_{\mathbf{s}_k}^\perp \mathbf{S}_k \right)^{-1} \mathbf{S}_k^* \mathbf{P}_{\mathbf{s}_k}^\perp \right) \\ &= \mathbf{E}_{\mathbf{s}_k \mathbf{S}_k}. \end{aligned} \quad \square$$

## APPENDIX B PROOF OF THEOREM VII.1

Let  $\mathbf{R}_{\alpha\alpha} = \mathbf{R}_{\alpha\alpha}^{1/2} \mathbf{R}_{\alpha\alpha}^{1/2}$  and  $\mathbf{R}_{\beta\beta} = \mathbf{R}_{\beta\beta}^{1/2} \mathbf{R}_{\beta\beta}^{1/2}$ , a permissible factorization for any positive definite matrix. Define  $\mathbf{F} = \mathbf{H} \mathbf{R}_{\alpha\alpha}^{1/2}$  and  $\mathbf{W} = \mathbf{S} \mathbf{R}_{\beta\beta}^{1/2}$  and rewrite  $\mathbf{R}$  as

$$\mathbf{R} = \mathbf{F} \mathbf{F}^* + \mathbf{W} \mathbf{W}^* + \sigma^2 \mathbf{I} = \mathbf{T} \mathbf{T}^* + \sigma^2 \mathbf{I} \quad (72)$$

where  $\mathbf{T} = [\mathbf{F} \ \mathbf{W}]$ . Then  $\mathbf{T}$  has the singular value decomposition  $\mathbf{T} = \mathbf{U}_T \mathbf{\Gamma} \mathbf{V}_T^*$ , where  $\mathbf{U}_T = \mathbf{U}_{\mathbf{H}\mathbf{S}}$  and  $\mathbf{\Gamma} = \mathbf{\Lambda}_{\mathbf{H}\mathbf{S}}$ . The pseudo-inverse of  $\mathbf{T}$  is given by  $\mathbf{T}^+ = \mathbf{V}_T \mathbf{\Lambda}_{\mathbf{H}\mathbf{S}}^{-1} \mathbf{U}_{\mathbf{H}\mathbf{S}}^*$  and from the definitions of  $\mathbf{G}$  and  $\mathbf{T}$  we see that  $\mathbf{G} = (\mathbf{T}^+)^* \mathbf{T}^+ \mathbf{H}$ . Since  $\mathbf{T}$  has full column rank, it is clear that

$$\mathbf{T}^+ \mathbf{H} = [\mathbf{F} \ \mathbf{W}]^+ \mathbf{F} \mathbf{R}_{\alpha\alpha}^{-1/2} = \begin{bmatrix} \mathbf{R}_{\alpha\alpha}^{-1/2} \\ \mathbf{0} \end{bmatrix}$$

and so

$$\mathbf{G} = (\mathbf{T}^+)^* \begin{bmatrix} \mathbf{R}_{\alpha\alpha}^{-1/2} \\ \mathbf{0} \end{bmatrix}.$$

We now employ Theorem II.2 to expand  $\mathbf{T}^+$

$$\mathbf{T}^+ = \begin{bmatrix} \mathbf{f}_1^* \mathbf{P}_{\mathbf{F}_1 \mathbf{W}}^\perp / \left( \mathbf{f}_1^* \mathbf{P}_{\mathbf{F}_1 \mathbf{W}}^\perp \mathbf{f}_1 \right) \\ \vdots \\ \mathbf{f}_r^* \mathbf{P}_{\mathbf{F}_r \mathbf{W}}^\perp / \left( \mathbf{f}_r^* \mathbf{P}_{\mathbf{F}_r \mathbf{W}}^\perp \mathbf{f}_r \right) \\ \mathbf{w}_1^* \mathbf{P}_{\mathbf{F} \mathbf{W}_1}^\perp / \left( \mathbf{w}_1^* \mathbf{P}_{\mathbf{F} \mathbf{W}_1}^\perp \mathbf{w}_1 \right) \\ \vdots \\ \mathbf{w}_{p-r}^* \mathbf{P}_{\mathbf{F} \mathbf{W}_{p-r}}^\perp / \left( \mathbf{w}_{p-r}^* \mathbf{P}_{\mathbf{F} \mathbf{W}_{p-r}}^\perp \mathbf{w}_{p-r} \right) \end{bmatrix}$$

where  $\mathbf{F}_i$  denotes the matrix  $\mathbf{F}$  with its  $i$ th column removed. Let  $\gamma_i = 1/(\mathbf{f}_i^* \mathbf{P}_{\mathbf{F}_i \mathbf{W}}^\perp \mathbf{f}_i)$  and  $\mathbf{a}_i = \gamma_i \mathbf{P}_{\mathbf{F}_i \mathbf{W}}^\perp \mathbf{f}_i$  for  $i = 1, \dots, r$ . Then we can write  $\mathbf{G}$  in the form

$$\mathbf{G} = (\mathbf{T}^+)^* \begin{bmatrix} \mathbf{R}_{\alpha\alpha}^{-1/2} \\ \mathbf{0} \end{bmatrix} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_r] \mathbf{R}_{\alpha\alpha}^{-1/2}.$$

It is clear that  $\text{rank}(\mathbf{G}) = r$  and that

- $\mathbf{P}_{\mathbf{G}} = \mathbf{P}_{[\mathbf{a}_1, \dots, \mathbf{a}_r]}$  (since  $\mathbf{R}_{\alpha\alpha}^{-1/2}$  has full rank);
- $\mathbf{P}_{\mathbf{W}}^\perp = \mathbf{P}_{\mathbf{S}}^\perp$  (since  $\mathbf{W} = \mathbf{S} \mathbf{R}_{\beta\beta}^{1/2}$  and  $\mathbf{R}_{\beta\beta}^{1/2} > 0$ );
- for each  $i$ ,  $\mathbf{a}_i = \gamma_i \mathbf{P}_{\mathbf{H}_i \mathbf{S}}^\perp \mathbf{f}_i \in \langle \mathbf{P}_{\mathbf{S}}^\perp \mathbf{F} \rangle$  (since  $\mathbf{P}_{\mathbf{H}_i \mathbf{S}}^\perp \mathbf{f}_i = \mathbf{P}_{\mathbf{S}}^\perp \mathbf{f}_i - \mathbf{P}_{\mathbf{P}_{\mathbf{S}}^\perp \mathbf{H}_i} \mathbf{f}_i$ , each in  $\langle \mathbf{P}_{\mathbf{S}}^\perp \mathbf{F} \rangle$ );
- and  $\langle \mathbf{P}_{\mathbf{S}}^\perp \mathbf{F} \rangle = \langle \mathbf{P}_{\mathbf{S}}^\perp \mathbf{H} \rangle$  since  $\mathbf{F} = \mathbf{H} \mathbf{R}_{\alpha\alpha}^{1/2}$  and  $\mathbf{R}_{\alpha\alpha}^{1/2}$  has full rank by assumption.

So  $\langle \mathbf{G} \rangle \subseteq \langle \mathbf{P}_{\mathbf{S}}^\perp \mathbf{H} \rangle$ , but  $\text{rank}(\mathbf{G}) = \text{rank}(\mathbf{P}_{\mathbf{S}}^\perp \mathbf{H})$  and the subspace spanned by a rank  $r$  matrix is identical to that spanned by any  $r$  linearly independent vectors in the subspace. Hence

$$\langle \mathbf{G} \rangle = \langle \mathbf{P}_{\mathbf{S}}^\perp \mathbf{H} \rangle \quad \text{and} \quad \mathbf{P}_{\mathbf{G}} = \mathbf{P}_{\mathbf{P}_{\mathbf{S}}^\perp \mathbf{H}}. \quad \square$$

## APPENDIX C PROOF OF LEMMA VII.2

By the Hoffman–Wielandt theorem [32] we have  $\|\hat{\mathbf{\Lambda}}_\Delta\|_2 \leq 2\delta$ , so the difficulty comes in bounding  $\|\hat{\mathbf{U}}_\Delta\|_2$ .

We begin by stating a lemma on eigenvalue sensitivity [9, Corollary 7.2.6].

*Lemma C.1:* Suppose  $\mathbf{R}, \mathbf{N} \in \mathbb{C}^{N \times N}$ , and that  $\mathbf{U} = [\mathbf{u}_1 \ \mathbf{u}_2]$  is unitary with  $\mathbf{u}_1 \in \mathbb{C}^N$ . Assume

$$\mathbf{U} \mathbf{R} \mathbf{U}^* = \begin{bmatrix} \lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix} \quad \text{and} \quad \mathbf{U}^* \mathbf{N} \mathbf{U} = \begin{bmatrix} \gamma_1 & \gamma_2^* \\ \gamma_2 & \mathbf{W} \end{bmatrix}$$

where  $\mathbf{D}$  is diagonal. If

$$\sigma(\lambda) = \min_i \{ \mathbf{D} - \lambda \mathbf{I} \}_{i,i} > 0$$

and  $\|\mathbf{N}\|_2 \leq \sigma/5$ , then there exists a  $\mathbf{p} \in \mathbb{C}^N$  with  $\|\mathbf{p}\|_2 \leq 4(\|\gamma_2\|_2/\sigma(\lambda))$  such that  $\hat{\mathbf{u}}_1 = (\mathbf{u}_1 + \mathbf{U}_2 \mathbf{p})/\sqrt{1 + \mathbf{p}^* \mathbf{p}}$  is a unit 2-norm eigenvector for  $\mathbf{R} + \mathbf{N}$ .

For our problem we see that  $\sigma(\lambda)$  defined in the preceding lemma is nonzero for each of the unique eigenvalues  $\lambda$  of  $\mathbf{\Lambda}_{\mathbf{H}\mathbf{S}}^2$ . In fact, for each  $\lambda_k$  associated with  $\mathbf{\Lambda}_{\mathbf{H}\mathbf{S}}^2$  we see that  $\sigma(\lambda_k) = \min_{j \neq k} |\lambda_k - \lambda_j|$ . We note further that  $\|\gamma_2\|_2 \leq \|\mathbf{N}\|_2$  by the properties of the 2-norm. Let  $\alpha_k = 4\|\gamma_2\|_2/\sigma(\lambda_k)$ , then we find that

$$\begin{aligned} \|\mathbf{u}_1 - \hat{\mathbf{u}}_1\| &= \|\mathbf{u}_1 - \hat{\mathbf{u}}_1\|_2 \\ &= \|\mathbf{u}_1(1 - 1/\sqrt{1 + \mathbf{p}^* \mathbf{p}}) - \mathbf{U}_2 \mathbf{p}/\sqrt{1 + \mathbf{p}^* \mathbf{p}}\| \\ &\leq \|\sqrt{1 - \mathbf{p}^* \mathbf{p}} - 1\| + \|\mathbf{U}_2 \mathbf{p}/\sqrt{1 + \mathbf{p}^* \mathbf{p}}\| \\ &\leq \sqrt{1 + \alpha_k^2} - 1 + \alpha_k. \end{aligned}$$

Since  $\mathbf{U}_{\mathbf{H}\mathbf{S}}$  are the eigenvectors associated with the  $p$  unique eigenvalues of  $\mathbf{R}$  corresponding to  $\mathbf{\Lambda}_{\mathbf{H}\mathbf{S}}^2$  we may bound the Frobenius norm of  $\mathbf{U}_{\mathbf{H}\mathbf{S}} - \hat{\mathbf{U}}_{\mathbf{H}\mathbf{S}}$  by

$$\begin{aligned} \|\mathbf{U}_{\mathbf{H}\mathbf{S}} - \hat{\mathbf{U}}_{\mathbf{H}\mathbf{S}}\|_F &= \left( \sum_{i=1}^p \|[U_{\mathbf{H}\mathbf{S}}]_i - [\hat{U}_{\mathbf{H}\mathbf{S}}]_i\|^2 \right)^{1/2} \\ &\leq \sqrt{p} \max_i \|[U_{\mathbf{H}\mathbf{S}}]_i - [\hat{U}_{\mathbf{H}\mathbf{S}}]_i\| \\ &\leq 4\sqrt{p} \max_{i \neq j} \left( \sqrt{1 + \left( \frac{\|\mathbf{N}\|_2}{|\lambda_i - \lambda_j|} \right)^2} - 1 + \frac{\|\mathbf{N}\|_2}{|\lambda_i - \lambda_j|} \right). \end{aligned}$$

Since

$$\|\mathbf{U}_{\mathbf{H}\mathbf{S}} - \hat{\mathbf{U}}_{\mathbf{H}\mathbf{S}}\|_2 \leq \|\mathbf{U}_{\mathbf{H}\mathbf{S}} - \hat{\mathbf{U}}_{\mathbf{H}\mathbf{S}}\|_F$$

the lemma follows.  $\square$

## REFERENCES

- [1] R. Lupas and S. Verdú, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 35, pp. 123–136, Jan. 1989.
- [2] A. Russ, "Noncoherent Detection for Nonlinear Binary Modulation in Gaussian CDMA Channels," M.S. Thesis, Friedrich Alexander Univ., Erlangen-Nuremberg, Germany, Sept. 1996.
- [3] M. McCloud, L. L. Scharf, and L. T. McWhorter, "Subspace coherence for detection in multiuser additive noise channels," in *Proc. SPAWC'97*, Paris, France, Apr. 1997, pp. 225–228.
- [4] M. Varanasi and A. Russ, "Noncoherent decorrelative multiuser detection for nonlinear nonorthogonal modulation," in *Proc. I.C.C. '97*, Montreal, Canada, June 1997.
- [5] M. McCloud and L. L. Scharf, "Generalized likelihood detection on multiple access channels," in *Proc. Asilomar'97*, Monterey, CA, Nov. 1997.
- [6] M. Varanasi and A. Russ, "Noncoherent decorrelative detection for nonorthogonal multipulse modulation over the multiuser Gaussian channel," *IEEE Trans. Commun.*, vol. 46, pp. 1675–1684, Dec. 1998.
- [7] X. Wang and H. V. Poor, "Blind multiuser detection: A subspace approach," *IEEE Trans. Inform. Theory*, vol. 44, pp. 677–689, Mar. 1998.
- [8] R. T. Behrens and L. L. Scharf, "Signal processing applications of oblique projection operators," *IEEE Trans. Signal Processing*, vol. 42, pp. 1413–1423, June 1994.
- [9] G. Golub and C. Van Loan, *Matrix Computations*. Baltimore, MD: Johns Hopkins Univ. Press, 1996.
- [10] M. Varanasi and T. Guess, "Achieving vertices of the capacity region of the Gaussian correlated-waveform multiple-access channel with decision feedback receivers," in *Proc. 1997 Int. Symp. Information Theory*, Ulm, Germany, June–July 1997, p. 270.
- [11] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Processing*, vol. 42, pp. 2146–2157, Aug. 1994.
- [12] L. L. Scharf, *Statistical Signal Processing*. Reading, MA: Addison-Wesley, 1991.
- [13] M. Varanasi, "Noncoherent detection in asynchronous multiuser channels," *IEEE Trans. Inform. Theory*, vol. 39, pp. 157–176, Jan. 1993.
- [14] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1980.
- [15] A. Björck and G. Golub, "Numerical methods for computing angles between linear subspaces," *Math. Comput.*, vol. 27, no. 123, pp. 579–594, July 1973.
- [16] M. Schwartz, W. Bennet, and S. Stein, *Communication Systems and Techniques*, New York: IEEE Press, 1996.
- [17] J. Proakis, *Digital Communications*, New York: McGraw-Hill, 1995.
- [18] R. Pawula, "Relations between the Rice  $I_0$ -function and the Marcum  $Q$ -function with applications to error rate calculations," *Electron. Lett.*, vol. 31, no. 24, pp. 2078–2080, Nov. 1995.
- [19] M. Loeve, *Probability Theory*, New York: Springer-Verlag, 1977.
- [20] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Processing*, vol. 39, pp. 1110–1121, May 1991.
- [21] L. Zhao, P. Krishnaiah, and Z. Bai, "On detection of the number of signals in presence of white noise," *J. Mult. Anal.*, vol. 20, pp. 1–25, Oct. 1986.
- [22] J. Bunch, C. Nielson, and D. Sorensen, "Rank-one modification of the symmetric eigenproblem," *Numer. Math.*, vol. 31, no. 1, pp. 31–48, 1978.
- [23] I. Karasalo, "Estimating the covariance matrix by signal subspace averaging," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, pp. 8–12, Jan. 1986.
- [24] R. DeGroat and R. Roberts, "Efficient, numerically stabilized rank-one eigenstructure updating," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 301–316, Feb. 1990.
- [25] P. Comon and G. Colub, "Tracking a few extreme singular values and vectors in signal processing," *Proc. IEEE*, vol. 78, pp. 1327–1343, Aug. 1990.
- [26] B. Champagne, "Adaptive eigendecomposition of data covariance matrices based on first-order perturbations," *IEEE Trans. Signal Processing*, vol. 42, pp. 2758–2770, Oct. 1994.
- [27] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Processing*, vol. 43, no. 1, pp. 95–107, Jan. 1995.
- [28] B. Yang, "An extension of the PASTD algorithm to both rank and subspace tracking," *IEEE Signal Processing Lett.*, vol. 2, no. 9, pp. 179–182, 1995.
- [29] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, pp. 387–392, Apr. 1985.
- [30] U. Madhow and M. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Commun.*, vol. 42, pp. 3178–3188, Dec. 1994.
- [31] M. McCloud and L. L. Scharf, Asymptotic analysis of the MMSE multiuser detector for nonorthogonal multipulse modulation, *IEEE Trans. Commun.*, Sept. 1999, submitted for publication.
- [32] R. Horn and C. Johnson, *Matrix Analysis*, New York: Cambridge Univ. Press, 1985.